SOME PROBLEMS OF ESTIMATION IN RESTRICTED PARAMETER SPACES

A Thesis Submitted
In Partial Fulfilment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

by

SOMESH KUMAR

to the

DEPARTMENT OF MATHEMATICS
INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

OCTOBER, 1988

To

My Parents

CERTIFICATE

This is to certify that the thesis 'Some Problems of Estimation in Restricted Parameter Spaces' submitted by Somesh Kumar in partial fulfilment of the degree of Doctor of Philosophy to the Department of Mathematics, I.I.T. Kanpur, is a record of bonafide research work carried out by him under my supervision and guidance. The results embodied in this thesis have not been submitted to any other University or Institute for the award of any diploma or degree.

D Sharma

(D. Sharma)
Professor
Department of Mathematics
Indian Institute of Technology
Kanpur.

October, 1988

ACKNOWLEDGEMENTS

It is with a deep sense of gratitude that I wish to thank Professor Divakar Sharma. His constant encouragement and valuable suggestions at hours of need have made this dissertation possible.

Sincere thanks are due to Vellaisamy, Tanuja, Kaushal, Bhaisaheb, Shubha, Neeraj and Rajiv for providing many cheerful moments and for bearing with me in difficult times. Special mention should be made of Neeraj, Anuja and Shubha for their help in the painstaking task of proofreading.

Finally, I appreciate very much the care and patience shown by Sri R.N. Srivastava in typing the thesis.

Somesh Kumar

Somesh Kumar

CONTENTS

LIST OF TABLES

CHAPTER - 1

Introduction and Summary

1.1 Introduction

The problem of estimation when the parameter space is restricted dates back to Hammersley (1950). He gave several examples of the situation where it is natural to restrict the parameter space to a subspace. Restriction of the parameter space reflects the prior information about the values of the parameter. Hammersley considered estimation of the means of normal and Poisson distributions when they are known to take only integer values. This type of problem arises, for example, in the estimation of the molecular weight of insulin where chemical theory restricts the unknown molecular weight to an integer. Katz (1961) considered estimation of $E_\omega(X)$ where X has a distribution with the density in the exponential family,

$$p_\omega(x) = \beta(\omega)\; e^{x\omega}, \qquad x \in \mathcal{X} \quad \text{and} \quad \omega \in T,$$

where $T = \{\omega: (\beta(\omega))^{-1} = \int_{\mathcal{X}} e^{x\omega}\, dx < \infty\}$.

When the parameter ω is known to be bounded below by a constant, say a, Katz considered the generalized Bayes estimate

$$\delta(x) = x + \beta(a)e^{ax} / \left(\int_a^\infty \beta(\omega)\; e^{x\omega}\, d\cancel{x}\right) \qquad d\omega$$

with respect to the uniform prior on $\{\omega: \omega \geq a\}$. He showed that $\delta(X)$ is admissible when the loss function is squared error. He

also proved its minimaxity when $p_{\omega}(x)$ is normal. A result, which gives sufficient condition for admissibility of an estimator with bounded risk, was given too. The condition is quite similar to Karlin's (1958) condition for admissibility.

Gupta and Rohatgi (1981) took up continuous distributions with unknown mean μ and unknown variance. They suggested several estimators for μ when it is known that μ lies in an interval (a,∞) (or $(-\infty,b)$). The estimators suggested lie completely in the range (a,∞) (or $(-\infty,b)$). The performances of these estimators were compared numerically for the normal, exponential and weighted difference of two independent chi-square distributions, weights being the unknown parameters. The last distribution occurs in balanced one-way variance components problem. It is the distribution of the usual unbiased estimator of the main effects variance component.

For a general density $f(x,\theta)$, where θ lies in a closed and compact space Ω, Ghosh (1964) considered the estimation of a bounded real function $g(\theta)$ when the loss function is

$$L(\theta,a) = |g(\theta)-a|^{p}, \quad p > 1 .$$

Restricting attention to the class $F_{\infty}^{(p)}$ of estimators with bounded risk, he proved that a unique minimax estimator exists provided $g(\theta)$ and $f(x,\theta)$ are continuous functions of θ. Under some further conditions on $f(x,\theta)$, it was shown that there exists a class C of estimators which is complete and every minimax sequence in C is a uniform minimax sequence. A sequence $\{\delta_n\}$ of estimators is said to be minimax if $\sup_{\theta \in \Omega} R(\theta,\delta_n) \to \sup_{\theta \in \Omega} R(\theta,\delta_o)$ as $n \to \infty$, where δ_o

is the unique minimax estimator. It is said to be a uniform minimax sequence if $R(\theta,\delta_n) \to R(\theta,\delta_o)$ uniformly in θ as $n \to \infty$.

In general the determination of minimax estimators is difficult. Let $\{v_1, v_2, \ldots\}$ be the basis of $F_\infty^{(p)}$. Ghosh showed that an approximation to the minimax estimator δ_o can be obtained by δ_n, the estimator minimax in the linear subspace spanned by n basis vectors $v_1, \ldots, v_n$. This method may be helpful, particularly in the problems which are slight variations of the standard problem. If δ^* is the conventional minimax estimator for the standard problem one can start off with a basis which has δ^* as the first element.

As an application of the above method, Ghosh considered the problem of estimating the mean of a random $N(\theta,1)$ variable, $\theta \in [-m,m]$. When θ is unrestricted, the sample mean $\bar{X}$ is the usual minimax estimator. Thus, for estimating θ restricted to $[-m,m]$, Ghosh started with basis $\bar{X}$, $\bar{X}^3$, $\bar{X}^5$ etc. and constructed a uniform minimax sequence. It turns out that the estimators obtained are too complicated to evaluate and the limiting minimax estimator cannot be found explicitly. However, the results indicate that the minimax estimator is Bayes with respect to a least favourable prior concentrated on a finite number of points. When the interval containing θ is small enough (for $m \le m_o = 1.05$ approximately), Casella and Strawderman (1981) showed that the prior is symmetric and puts all its probability on $\pm m$. For somewhat wider intervals they obtained sufficient conditions for the minimaxity of an estimator which is Bayes with respect to a three point prior. These conditions are satisfied for $m \in (1.4, 1.6)$. However, numerical evidence suggests that the conditions will be satisfied for

$m \in (m_o, 2)$. As m increases, the nature of the least favourable prior becomes complicated and numerical techniques are necessary to determine the support points and the probabilities assigned to them.

Bickel (1981) studied the behaviour of the least favourable prior distribution and the minimax risk for large m. Using an identity of Brown (1971), he exhibited that the minimax risk can be obtained by minimizing the Fisher information among prior distributions concentrated on $[-m,m]$. Earlier, it was shown by Huber (1974) that the distribution G_1 with density

$$g_1(y) = \begin{cases} \cos^2(\frac{\pi}{2} y), & |y| \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

minimizes Fisher information among distributions on $[-1,1]$. Bickel rescaled the least favourable priors to $[-1,1]$ and showed that they converge weakly to G_1 as $m \to \infty$. Thus, G_1 is the approximate least favourable prior. The minimax risk is $1 - \frac{\pi^2}{m^2} + o(\frac{1}{m^2})$ as $m \to \infty$.

Gatsonis, MacGibbon and Strawderman (1987) numerically compared performances of the Bayes estimator δ_m with respect to the uniform prior on $[-m,m]$, the maximum likelihood estimator, the Bayes estimator with respect to Bickel's prior $\frac{1}{m}\cos^2(\frac{\pi}{2m} y)$, $|y| \le m$ and some other estimators. They concluded that δ_m is a good choice among these.

Melkman and Ritov (1987) generalized the results of Bickel to a distribution with the location parameter density. They showed that the only conditions required are $EX = \theta$ and $EX^4 < \infty$.

Bickel (1981) also considered the estimation of the mean of a p-variate normal distribution ($p \ge 2$) when the mean vector is

restricted to take values inside a sphere. He showed that the results for $p = 1$ can be extended to $p \geq 2$.

Moors (1985) discussed truncated estimation problems invariant under a finite group G with the induced group $\tilde{G}$ a subgroup of the group of linear transformations. These problems are called linvariant by Moors. When the loss function is quadratic, under mild conditions he obtained a subspace A_x of the action space for each observed x such that any linvariant estimator taking values outside A_x with positive probability would be inadmissible. The dominating estimator was provided and various applications of the result were given. A striking consequence of this result is that in truncated estimation problems, linvariant estimators taking values on or near the boundary of the parameter space would be inadmissible. Moors, in particular, studied the problem of estimating binomial parameter θ, when θ is restricted to an interval $[1-P, P]$ for some known $P \in (\frac{1}{2}, 1]$. This problem has its origin in randomized response models, where the design of the experiment itself leads to a restriction of the parameter space (see, for exampl Horvitz et al. (1976)). Moors obtained minimax estimators and the least favourable priors numerically for specific values of P and for n = 3(1)16 when the loss function is quadratic and for n = 3(1)11 when the loss function is weighted quadratic.

Estimation of binomial parameter θ, when the parameter space is a subset of $[0, 1]$, has been considered before. Skibinsky and Cote (1963) investigated the case when the prior knowledge implies that θ is outside the interval $[1-P, P]$ with a given small probability. They showed that the sample proportion is inadmissible.

Blum and Rosenblatt (1967) studied $\mathcal{F}$-minimaxity when $\theta \in [0,P]$. Rafsky (1976) noted that for finite populations, the parameter space is, in fact, discrete. Using this fact he obtained an estimator better than the sample proportion. Schafer (1976) offered some alternative estimators to the usual minimax estimator, when $\theta \in (0,1)$. This problem is encountered in estimation of tail areas in an unspecified positive probability distribution function on $(0,\infty)$.

Next, we discuss in detail two special problems in restricted parameter spaces, which have received attention of many researchers. These are the estimation of ordered parameters when the ordering is known, and the estimation of the common location of several populations.

1.1.1 A Review of the Problem of Estimating Ordered Parameters

The problem of estimating ordered parameters when the ordering is not known is widely discussed in literature (see Dudewicz and Koo (1982) for a detailed bibliography). However, not much attention has been paid to the case when the ordering is known. This is, clearly, a problem of estimation in restricted parameter space and has its origin in various agricultural, industrial and economic experiments. Suppose we want to estimate the average yields θ_1 and θ_2 under treatments 1 and 2 respectively; treatment 1 is using a certain fertilizer for the crop while treatment 2 is not using any fertilizer. Then it is possible to assert a priori that $\theta_1 \geq \theta_2$. In estimating the average incomes say θ_i, $i = 1,\ldots,4$ of four classes of employees in a certain

establishment, it is quite reasonable to assume an ordering among θ_i's according to the grade of employees. In the development of a system, engineering changes are made in stages to correct design deficiencies and thereby to increase reliability. Thus, if we have k stages in which the changes are made, then at each stage we expect the average production to increase. If θ_i is the average production at the i^{th} stage, we may assume that $\theta_1 < \ldots < \theta_k$.

Most of the literature on estimating ordered parameters is concerned with maximum likelihood estimation for specific distributions. Barlow, Bartholomew, Bremner and Brunk (1972) give detailed description of the results in this area. (~~Recently~~) Sackrowitz and Strawderman (1974) considered admissibility question of the maximum likelihood estimators (MLE's) of the parameters of k binomial populations, when the parameters have a known ordering. Suppose $X_1,\ldots,X_k$ are k independent random variables with X_i having a binomial distribution with parameters θ_i and n_i, $i = 1,\ldots,k$. Then, for estimating $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ when $\theta_1 \leq \ldots \leq \theta_k$, with the loss function the sum of strictly convex losses, Sackrowitz and Strawderman showed that the MLE is inadmissible under the following conditions:

(i) If $k = 2$, $n_1 \neq 1$, $n_2 \neq 1$ and $n_1+n_2 \geq 7$,

(ii) If $k = 3$, $(n_1,n_3) \neq (1,1)$ and $n_1+n_2+n_3 \geq 7$,

(iii) If $k > 3$, $n_1 + n_2 + \ldots + n_k \geq 7$.

The inadmissibility of the MLE was proved by showing that it is not admissible in a suitably chosen subproblem and this fact ? was proved by showing that the MLE cannot be Bayes in that subproblem. However, this method of proving inadmissibility does not yield a better estimator. Sackrowitz (1982) presented a method of

constructing, in the cases where the MLE is inadmissible, an estimator which is better than the MLE. The loss function was taken to be the sum of squared errors. The argument used in obtaining the improvement is the following: Consider a nonempty subset S^* of the sample space S and let $\underline{X}^*$ be the restriction of $\underline{X} = (X_1,...,X_k)$ to S^* in the sense that $P_{\underline{\theta}}(\underline{X}^* = \underline{x}) = P_{\underline{\theta}}(\underline{X} = \underline{x}/\underline{X} \in S^*)$. It can be easily proved that any estimator of $\underline{\theta}$ inadmissible in the set up$(\underline{X}^*, S^*)$ is also inadmissible in the original set up $(\underline{X}, S)$. Now the subsets of S where the restricted MLE is inadmissible are identified and a better estimator is obtained by altering the MLE on these subsets. However, there are practical problems in this approach. A general expression for the new estimator cannot be given and as there is no unique ordering on the procedure for identifying sets where the MLE is inadmissible, the method results in a number of different estimators, each one better than the MLE. Computational difficulties are faced while determining the amount of improvement. Numerical comparisons for a few simple cases of $k = 2$ show a marginal reduction in the risk over a very large portion of the parameter space.

For the problem of simultaneous estimation of two ordered binomial parameters, Katz (1963) suggested a mixed estimator which improves an estimator $\underline{\delta} = (\delta_1, \delta_2)$, not satisfying the order relation $\delta_1 \leq \delta_2$ with positive probability. The mixed estimator $\underline{\delta}_{\alpha^+}$ for $\underline{\delta}$ is

$$\underline{\delta}_{\alpha^+} = (\alpha\delta_1 + (1-\alpha)\delta_2,\ (1-\alpha)\delta_1 + \alpha\delta_2)\ ,$$

where $\alpha = 1$, if $\delta_1 \leq \delta_2$, and
$= \alpha^+$, if $\delta_1 > \delta_2$, $0 < \alpha^+ < 1$.

Katz obtained estimators minimax in the class $\{\underline{\delta}_{\alpha^+}: 0 \leq \alpha^+ \leq 1\}$. He also considered the problem of simultaneous estimation of two ordered normal means θ_1 and θ_2; $\theta_1 \leq \theta_2$, when the variances are the same and known. He proved admissibility and minimaxity of the generalized Bayes estimator of (θ_1, θ_2) with respect to the uniform prior on the space $\theta_1 \leq \theta_2$. However, as Blumenthal and Cohen (1968b) mention, his proofs are inadequate.

Blumenthal and Cohen (op. cit.) considered the estimation of location parameters θ_1 and θ_2 $(\theta_1 \leq \theta_2)$ from two continuous populations, when the loss function is the sum of squared errors. They developed sufficient conditions for the minimaxity and admissibility of the generalized Bayes estimator $\underline{\delta}_p$ of $\underline{\theta}$ with respect to the uniform prior on the space $\theta_1 \leq \theta_2$.

Sackrowitz (1970) considered independent discrete random variables $X_1, X_2, \ldots$ with probability functions $f(x_1,\theta_1), f(x_2,\theta_2), \ldots$ respectively. Estimation of θ_k on the basis of $(X_1,\ldots,X_k)$, when $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$, is taken up. Sackrowitz showed that the minimax value M_k is the same for all $k = 1,2,\ldots$. If $t(X_1,\ldots,X_s)$ is an admissible estimator for θ_s then the estimator $t(X_{k+1},\ldots,X_{k+s})$ is admissible for θ_{k+s}. Hence if $t(X_1)$ is admissible for θ_1 with the constant risk then $t(X_s)$ will be the unique admissible minimax estimator of θ_s among those based on $(X_1,\ldots,X_s)$. Sackrowitz also obtained a sequence of estimators which is asymptotically subminimax (as $s \to \infty$). A subminimax estimator is one which has maximum risk slightly larger than the minimax risk but the risk

function substantially below that of a minimax estimator over a large portion of the parameter space.

Cohen and Sackrowitz (1970) considered the above estimation problem for continuous random variables with location parameter densities. When the variables are normal with known unequal variances and $k = 2$, they showed that an estimator based on X_2 alone with bounded risk is inadmissible. They also considered the generalized Bayes estimator of θ_2 with respect to the uniform prior on $\theta_1 \leq \theta_2$ and proved it to be admissible and minimax. For a general k they considered normal populations with equal known variances and obtained a class of admissible estimators based on X_k alone. For symmetric location parameter densities if $P_{\underline{\theta}}(X_1-X_2 > 0) > 0$ for some $\underline{\theta}$, $\theta_1 \leq \theta_2$ then X_2 was proved to be inadmissible. A similar result for confidence intervals was obtained.

1.1.2 A Review of the Problem of Estimation of Common Location

Let $\underline{X} = (X_1,\ldots,X_m)$ and $\underline{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from $N(\mu,\sigma_1^2)$ and $N(\mu,\sigma_2^2)$ populations respectively. The variances σ_1^2 and σ_2^2 are unequal and unknown. We estimate μ when the loss function is

$$L_1(\hat{\mu},\mu) = (\hat{\mu}-\mu)^2$$

or $$L_2(\hat{\mu},\mu) = (\hat{\mu}-\mu)^2/\sigma_1^2 .$$

This problem of estimating the common mean of two normal populations has its origin in the problem of recovery of interblock information in balanced incomplete block designs. Here two independent unbiased estimators (intrablock and interblock) for the treatment contrast are available. The aim is to combine the two estimators to get a

better estimator. The problem of recovery of interblock information has received considerable attention (see Kubokawa (1988) for a bibliography). Various combined estimators for μ have been suggested. One of the most commonly used estimators of μ was given by Graybill and Deal (1959),

$\hat{\mu}_{GD} = \dfrac{S_2\bar{X} + S_1\bar{Y}}{S_1 + S_2}$, $\bar{X}$, $\bar{Y}$, S_1 and S_2 being the sample means and sample sum of squares. Graybill and Deal showed that $\hat{\mu}_{GD}$ is unbiased and dominates both $\bar{X}$ and $\bar{Y}$ if and only if both the sample sizes are at least 11.

When both m and n are less than 11, $\hat{\mu}_{GD}$ does not improve $\bar{X}$ or $\bar{Y}$. For this case, Zacks (1966) proposed two classes of testimators:

$$
\begin{aligned}
\hat{\mu}(\rho^*) &= \frac{\bar{X}+\bar{Y}}{2}, \quad \text{if } \frac{1}{\rho^*} \le \frac{S_2}{S_1} \le \rho^*, \quad 1 \le \rho^* \le \infty \\
&= \hat{\mu}_{GD} \qquad \text{otherwise}
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{\mu}(\rho^*) &= \bar{Y}, \quad \text{if } \frac{S_2}{S_1} < \frac{1}{\rho^*} \\
&= \frac{\bar{X}+\bar{Y}}{2}, \quad \text{if } \frac{1}{\rho^*} \le \frac{S_2}{S_1} \le \rho^* \\
&= \bar{X}, \quad \text{if } \frac{S_2}{S_1} > \rho^*, \quad 1 \le \rho^* \le \infty.
\end{aligned}
$$

When the samples are of equal size 3, Zacks obtained explicit formulae of the efficiency functions of $\hat{\mu}(\rho^*)$ and $\tilde{\mu}(\rho^*)$, efficiency of an estimator μ^* is defined as $\mathrm{Var}(\hat{\mu}_o)/\mathrm{Var}(\mu^*)$, where $\hat{\mu}_o = (\sigma_2^2 \bar{X} + \sigma_1^2 \bar{Y})/(\sigma_1^2 + \sigma_2^2)$. The expressions of the efficiency function are quite complicated and numerical calculations are carried out. It is observed that the estimator $\hat{\mu}(\rho^*)$ dominates $\tilde{\mu}(\rho^*)$ in

$\{\rho: \frac{1}{6} \leq \rho \leq 6\}$, where $\rho = \sigma_2^2/\sigma_1^2$, for all values of ρ^*. For ρ^* close to one, $\hat{\mu}$ is uniformly better than $\tilde{\mu}$ and the efficiency of $\hat{\mu}$ exceeds that of $\tilde{\mu}$ by about 25% in the neighbourhood of $\rho = 1$. Zacks recommends the estimator $\hat{\mu}(\rho^*)$.

The case, when the samples are of equal size 5, was found to be of no special interest. In fact, the efficiency function of $\hat{\mu}(1)$ itself is very complicated. It was observed numerically that the gain in efficiency over the range $1 \leq \rho \leq 6$ is very slight.

Mehta and Gurland (1969) investigated the admissibility of $\hat{\mu}_{GD}$ when the samples are of equal size and $\rho \geq 1$. They considered estimators of the form

$$\hat{\mu}(\varphi) = \varphi(F)\bar{X} + (1-\varphi(F))\bar{Y} , \qquad F = S_2/S_1$$

and proposed estimators

(i) $T_1 = \hat{\mu}(\varphi)$, where

$$\varphi(F) = \frac{C+F}{C+A+F} ;$$

(ii) $T_2 = \hat{\mu}(\varphi)$, where

$$\varphi(F) = 1/2 , \quad \text{if } F < k^o$$

$$= \frac{F}{F+A} , \quad \text{if } F \geq k^o ;$$

(iii) $T_3 = \hat{\mu}(\varphi)$, where

$$\varphi(F) = \frac{(C+F)^{1/2}}{(C+F)^{1/2}+A}$$

with A, C, k^o specific constants to be chosen. It was shown that (i) corresponding to each n, there exists a T_1 which dominates $\hat{\mu}_{GD}$ (ii) the best choice of T_2, for a specified A, is given by $k^o = A$ (iii) for n=3, the limiting variance of T_3 becomes infinite as $\rho \to \infty$.

The efficiencies of T_1, T_2 and T_3 are compared numerically with that of $\hat{\mu}_{GD}$ for specific choices of A, C, k^o and ρ when n = 3. The estimators T_1 and T_2 are found to be more efficient than $\hat{\mu}_{GD}$ for $\rho \geq 1$. For $1 \leq \rho \leq 54.6$, T_3 is more efficient than $\hat{\mu}_{GD}$. Although T_1, T_2 and T_3 perform better than $\hat{\mu}_{GD}$, T_1 is the best, in view of the fact, that it has higher efficiency than that of $\hat{\mu}_{GD}$ for all $\rho \geq 1$. If we are interested in a particular interval of ρ values, say $1 < \rho < 20$, then T_3 is the best.

For n = 5, 7, 9 and 11 and ρ in the interval $[1, 54.6)$ it is exhibited that T_1 with A = 1 and C =.4 performs better than $\hat{\mu}_{GD}$.

For unequal sample sizes, Nair (1982) showed that the estimator T_2 with $A = k^o = 1$ is uniformly better than $\hat{\mu}_{GD}$ when $\sigma_2^2 > \sigma_1^2$ and gave an upperbound for the improvement in the risk.

Zacks (1970) considered estimators equivariant under the transformations $X_i \to \alpha + \beta X_i$, $Y_i \to \alpha + \beta Y_i$, $i = 1,\ldots,n$, $-\infty < \alpha < \infty$, $\beta \neq 0$. It was shown that these estimators are unbiased and also an expression for Bayes equivariant estimators was obtained. It was shown that $\bar{X}$ is a minimax estimator, when the loss is L_2. Zacks also gave the form of fiducial equivariant estimators and showed that Bayes equivariant and fiducial equivariant estimators are admissible.

Cohen and Sackrowitz (1974) considered the samples of equal size and obtained classes of minimax estimators when the loss is L_2. These estimators dominate $\bar{X}$ for $n \geq 6$ (see Wijsman (1976) for the correction). Using a symmetry argument it was shown that the estimator

$$\hat{\mu}_{CS} = (1-G^*(Z))\overline{X} + G^*(Z)\overline{Y} \ ,$$

where
$$G^*(Z) = F(1, (\frac{3-n}{2}), (\frac{n-1}{2}), z) \qquad \text{for } 0 \leq z \leq 1$$
$$= \frac{n-3}{n-1}\,\frac{1}{z}\,F(1, (\frac{5-n}{2}), (\frac{n+1}{2}), \frac{1}{z}) \qquad \text{for } z \geq 1,$$

and F is the hypergeometric function; is better than both $\overline{X}$ and $\overline{Y}$ for $n \geq 10$. Cohen and Sackrowitz also gave confidence intervals which dominate $(\overline{X}-h, \overline{X}+h)$.

Brown and Cohen (1974) obtained estimators better than $\overline{X}$ if the first sample size is greater than 1 and the second sample size is greater than 2. Some of these estimators can be extended to dominate $\overline{X}$ when we have more than two populations. A class of estimators dominating $\overline{X}$ was obtained also by Khatri and Shah (1974).

Norwood and Hinkelmann (1977) considered estimation of the common mean μ of k normal populations with unknown and unequal variances $\sigma_1^2,\ldots,\sigma_k^2$. Let $\overline{X}_i$ and S_i be the mean and sum of squares respectively from the i^{th} sample. Norwood and Hinkelmann showed that an extension of the Graybill Deal estimator $\hat{\mu}_{GD}$,

$$\hat{\mu}_{NH} = \frac{\sum_{i=1}^{k} \overline{X}_i/S_i}{\sum_{i=1}^{k} 1/S_i} \ ,$$

is better than each $\overline{X}_i$, if each sample size is more than 10 or, if one sample size is 10 and all others are more than 18.

Shinozaki (1978) and Bhattacharyya (1979) proved that the estimator

$$\hat{\mu}_{SB} = \frac{\sum_{i=1}^{k} C_i S_i^{-1} \bar{X}_i}{\sum_{i=1}^{k} C_i S_i^{-1}} ,$$

where C_i's are constants, is better than each $\bar{X}_i$ if and only if

$$\frac{(m_k+1)}{2(m_i-5)} \le \frac{C_k}{C_i} \le \frac{2(m_k-5)}{(m_i+1)} , \qquad i = 1,\ldots,k-1,$$

with m_i as the i^{th} sample size.

Bhattacharyya (1984) showed that the left part of the above inequality is unnecessary. For a proof of this fact he developed two general inequalities.

Sinha and Mouqadem (1982) investigated admissibility of the Graybill-Deal estimator $\hat{\mu}_{GD}$ when k = 2. They considered four classes of unbiased estimators

$$\mathcal{C} = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \le \Psi(S_1, S_2, D^2) \le 1\} ,$$

$$\mathcal{C}_o = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \le \Psi(S_2/S_1) \le 1\} ,$$

$$\mathcal{C}_1 = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \le \Psi(S_1, S_2) \le 1\} \text{ and}$$

$$\mathcal{C}_2 = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \le \Psi(S_1(\bar{Y}-\bar{X})^{-2},\ S_2(\bar{Y}-\bar{X})^{-2}) \le 1\} .$$

It was shown that $\hat{\mu}_{GD}$ is admissible in $\mathcal{C}_o$, extended admissible in $\mathcal{C}$ but neither Bayes nor limiting Bayes in $\mathcal{C}_2$. Some estimators admissible in classes $\mathcal{C}_1$ and $\mathcal{C}_2$ were given.

Bhattacharyya (1980), Kubokawa (1987a, 1987b) have also obtained classes of estimators dominating $\bar{X}$. Kubokawa (1987b) develops admissible estimators from these classes of estimators. Kubokawa (1987c) has generalized the result of Bhattacharyya (1984) to a class of more general loss functions $W(|\hat{\mu}-\mu|^r)$, where W is a concave function and $0 < r \le 2$.

Shinozaki (1978) has also considered the estimation of the common mean vector of k p-variate normal populations. He gave a class of unbiased estimators which, for any p, dominate each of the k sample means. However, these unbiased estimators can also be improved by biased estimators for $p \geq 3$.

Hogg (1960b) was the first one to consider the general problem of estimating the common location parameter of several populations which have the same form of the density but different scale parameters. He suggested a method of constructing unbiased estimators. The method used the fact that if the density under consideration is symmetric about the location, the conditional expectation of an odd location-scale statistic given an even location free-scale statistic equals the location parameter. A statisti is said to be odd location-scale if it is equivariant under affine group of transformations $\{g: gx = \alpha x+\beta, \alpha \text{ real and } \beta \neq 0\}$. An even location free-scale statistic is one invariant under translation and equivariant under scalar multiplication. Combining odd location-scale estimators from the samples with the weights as functions of even location free-scale statistics, he obtained an unbiased estimator of the common location parameter.

Cohen (1976) gave examples of where the problem of estimating a common location for two populations could arise. He obtained a combined estimator δ_a which is unbiased and which, for $0 < a \leq a^*(m,n)$; m,n the sizes of the sample from two populations, improves upon the first sample estimator Ψ_x under mild conditions.

Bhattacharyya (1981) showed that Cohen's estimator δ_a improves upon the estimator Ψ_x, if $0 < a \leq A(m,n)$. The bound $A(m,n)$

is shown to be larger than $a^*(m,n)$ and easier to calculate. In fa
for uniform distribution tabulated values show that $a^*(m,n)$ is ne
larger than 1 for both the pairs (m,n) and (n,m) whereas $A(m,n) \geq$
for both the pairs (m,n) and (n,m) for $m = n \geq 25$ and for $n \geq 35$
whenever $m \leq n+5$. As will be seen later in Chapter 6, if the boun
is greater than or equal to 1 for both the pairs (m,n) and (n,m),
the estimator δ_1 is better than both Ψ_X and Ψ_Y, the estimators fro
the two samples.

Akai (1982) modified the estimator δ_a and obtained sufficient conditions for it to dominate Ψ_X and Ψ_Y.

So far in this section, we have restricted attention to th
case, where the samples come from independent populations. Halper
(1961) was probably the first one to consider the estimation of th
common mean of a general multivariate normal distribution. He obta
ined the maximum likelihood estimator (MLE) and showed that its variance is equal to that of the minimum variance linear unbiased estimator except for a multiplicative factor which does not depend on the parameters and approaches unity as the sample size becomes large. It is interesting to note that when the correlations are assumed to be zero, that is, in the case of independent populations the MLE cannot be obtained explicitly and so has not been considere by authors.

Rastogi and Rohatgi (1974) considered the above estimation problem for a bivariate normal distribution. Let $(X_1,Y_1),\ldots(X_n,Y_n$
be a random sample from a bivariate normal population with common mean μ, $\mathrm{Var}(X_i) = \sigma_1^2$, $\mathrm{Var}(Y_i) = \sigma_2^2$, $\sigma_1^2 \neq \sigma_2^2$ and the correlation coefficient ρ. They evaluated the variance expressions of the two estimators $\hat{\mu}_1$ and $\hat{\mu}_2$, where

is the average of two sample means and

$$\hat{\mu}_2 = \bar{Y} + (\bar{X}-\bar{Y}) \frac{(S_{22}-S_{12})}{(S_{11}+S_{22}-2S_{12})}$$

is the MLE, S_{11}, S_{12} and S_{22} being the components of the sample covariance matrix. Numerical computations show that as n becomes large the ratio $Var(\hat{\mu}_1)/Var(\hat{\mu}_2)$ tends to be larger than 1 except for $\sigma_1^2 = \sigma_2^2$. In general, for fixed ρ, the ratio $Var(\hat{\mu}_1)/Var(\hat{\mu}_2)$ decreases as σ_2^2/σ_1^2 increases to one and for fixed σ_2^2/σ_1^2, it increases as ρ increases in absolute value. Rastogi and Rohatgi also considered the case when additional N-n observations are available on X variable. Such a situation arises, for example, when two measuring devices are used to obtain readings on each of a number of objects or units and one of the measuring devices stops functioning after taking some readings. Two alternative estimators were suggested for this case and their performances compared numerically with that of $\hat{\mu}_2$. One of the estimators Z is a convex combination of $\hat{\mu}_2$ and the average of the last (N-n) observations on X with weights n/N and (N-n)/N. Z is seen to perform better than $\hat{\mu}_2$ in most part of the parameter space.

Recently Krishnamoorthy and Rohatgi (1988a) have proposed estimators $\hat{\mu}_3$ and $\hat{\mu}_4$ (for the common mean of a bivariate normal distribution) and compared their risk performances numerically with that of $\hat{\mu}_2$. $\hat{\mu}_3$ is seen to perform satisfactorily.

Glesser (1987a, 1987b) and Krishnamoorthy and Rohatgi (1988b have considered the problem of estimating the mean response when the control variables are used. By making a linear transformation this problem is seen to be equivalent to that of estimating the

common mean. Glesser obtained the MLE and proposed a class of unbiased estimators dominating it. Krishnamoorthy and Rohatgi (1988b) have suggested another unbiased estimator which dominates the MLE over a large portion of the parameter space.

In this thesis we take up some of the problems discussed above. A summary of the results obtained is given in next section.

1.2 Summary of the Results in the Thesis

In Chapters 3 and 4, we take up the problem, discussed in Section 1.1.1, of estimating ordered parameters when the ordering is known. Let $(X_{i1},\ldots,X_{in})$ be a random sample from a continuous population with the density $f_i(x-\theta_i)$, $i = 1,\ldots,k$. It is assumed that the k samples are drawn independently. We take

$$\int_{-\infty}^{\infty} x\, f_i(x)\, dx = 0\ , \qquad i = 1,\ldots,k$$

without loss of generality. Also, let θ_i's be ordered, say $\theta_1 \le \theta_2 \le \cdots \le \theta_k$.

In Chapter 3, the problem of simultaneous estimation of $\underline{\theta} = (\theta_1,\ldots,\theta_k)$, when the loss function is squared error, that is,

$$L_1(\underline{\theta},\underline{a}) = |\underline{\theta}-\underline{a}|^2 = \sum_{i=1}^{k} (\theta_i-a_i)^2 \ ; \tag{1.2.1}$$

is considered. When the parameter space is R^k, the most commonly used estimator of $\underline{\theta}$ is the Pitman estimator $\underline{X} = (X_1,\ldots,X_k)$, where

$$X_i = \frac{\int_{-\infty}^{\infty} \theta_i \prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_i}{\int_{-\infty}^{\infty} \prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_i}\ , \qquad i = 1,\ldots,k\ . \tag{1.2.2}$$

For θ_i's ordered, an estimator of interest is $\underline{\delta}_p$, the analogue of the Pitman estimator $\underline{X}$. It is defined to be the

generalized Bayes estimator of $\underline{\theta}$ with respect to the uniform prior on the space $\{\underline{\theta}: \theta_1 \leq \ldots \leq \theta_k\}$. If the components of $\underline{\delta}_p$ are denoted by $\delta_{p1},\ldots,\delta_{pk}$ respectively, we have

$$\delta_{pr}(X_{11},\ldots,X_{kn}) = \frac{\int\limits_{\theta_1\leq\ldots\leq\theta_k}\cdots\int \theta_r \prod_{i=1}^{k}\prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_1\ldots d\theta_k}{\int\limits_{\theta_1\leq\ldots\leq\theta_k}\cdots\int \prod_{i=1}^{k}\prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_1\ldots d\theta_k},$$

$$r = 1,\ldots,k . \qquad (1.2.3)$$

Hereafter, $\underline{\delta}_p$ will be simply called the Pitman estimator of $\underline{\theta}$.

Blumenthal and Cohen (1968b) considered $k = 2$ and identical f_i's. When the parameter space is any subset Σ_o of R^2, not necessarily $\{\underline{\theta}: \theta_1 \leq \theta_2\}$, they proved under a mild condition on Σ_o that any estimator, with the risk function not larger than the constant risk of the estimator $\underline{X}$ for all parameter values in Σ_o, is minimax for $\underline{\theta}$, $\underline{\theta} \in \Sigma_o$. In particular, the condition is satisfied when the parameter space is $\{\underline{\theta}: \theta_1 \leq \theta_2\}$. We, in Section 3.2, extend this result to a general k and nonidentical f_i's. A simple corollary of this result is that the usual estimator $\underline{X}$ remains minimax when $\underline{\theta}$ is restricted to a subspace of R^k. We also give a similar result for the problem of estimating a component θ_i of $\underline{\theta}$ when the loss function is squared error.

In Section 3.3, we take $k = 2$ and f_i as $N(\theta_i, 1)$, $i = 1,2$; $\theta_1 \leq \theta_2$. The properties of the mixed estimator $\underline{\delta}_{\alpha^+}(\underline{X})$, given by

$$\delta_{\alpha^+}(\underline{X}) = (\alpha X_1 + (1-\alpha)X_2 , (1-\alpha)X_1 + \alpha X_2) ;$$

where $\alpha = 1$ when $X_1 \leq X_2$

$= \alpha^+$ when $X_1 > X_2$, α^+ a constant,

are investigated. We show that $\underline{\delta}_{\alpha^+}(\underline{X})$, $0 \le \alpha^+ < 1$, dominates $\underline{X}$ and so is minimax. A class of estimators admissible among the mixed estimators is obtained. Another class of minimax estimators is also given. It is proved that the mixed estimator $\underline{\delta}_{1/2}(\underline{X})$ is the best choice in this class. We, however, show that any mixed estimator is inadmissible. The proof uses the fact that the class of generalized Bayes estimators is complete. We prove that $\underline{\delta}_{\alpha^+}$ is not generalized Bayes and so is inadmissible. However, this method of proving inadmissibility does not yield a better estimator. We do not know what estimators will improve $\underline{\delta}_{\alpha^+}$. Finally, a comparison between the Pitman estimator $\underline{\delta}_p$ and the mixed estimator $\underline{\delta}_{1/2}$ is made.

In Section 3.4, we study the case of two normal populations with unequal but known variances. As before, the MLE and the mixed estimators $\underline{\delta}_{\alpha^+}(\underline{X})$, $0 \le \alpha^+ < 1$, dominate $\underline{X}$ and thus, are minimax. However, unlike the equal variances case, the MLE, here, improves some of the mixed estimators including $\underline{\delta}_{1/2}(\underline{X})$. The Pitman estimator $\underline{\delta}_p$, as mentioned by Cohen and Sackrowitz (1970), is admissible. We prove that it is minimax. We also compare the risk function of $\underline{\delta}_p$ with that of the MLE and of the mixed estimator $\underline{\delta}_{1/2}$ and observe that $\underline{\delta}_p$ does not improve any.

The problem of Section 3.4 is a special case of the densities $\frac{1}{\sigma_1} f(\frac{x_1-\theta_1}{\sigma_1})$ and $\frac{1}{\sigma_2} f(\frac{x_2-\theta_2}{\sigma_2})$, $\theta_1 \le \theta_2$, with σ_1 and σ_2 unequal but known. In Section 3.5, we consider a still more general problem of estimating location parameters θ_1 and θ_2 of densities $f_1(x_1-\theta_1)$ and $f_2(x_2-\theta_2)$, $\theta_1 \le \theta_2$; where f_i's may be different, and obtain sufficient conditions for the minimaxity of the Pitman estimator $\underline{\delta}_p$ defined by (1.2.3). This is a generalization

of the results of Blumenthal and Cohen (1968b) who considered densities $f(x_1-\theta_1)$ and $f(x_2-\theta_2)$, $\theta_1 \leq \theta_2$. One of the sufficient conditions for minimaxity is the monotonicity of certain ratios of the distribution function and the corresponding density. It is seen to be satisfied for normal, uniform and gamma distributions.

The results of Chapter 3 are due to appear in Kumar and Sharma (1988) under the title "Simultaneous Estimation of Ordered Parameters".

In Chapter 4, we consider k independent normal random variables $X_1,\ldots,X_k$ with means $\theta_1,\ldots,\theta_k$ respectively, $\theta_1 \leq \ldots \leq \theta_k$ and the common variance unity. For estimating $\underline{\theta} = (\theta_1,\ldots,\theta_k)$, when the loss function is (1.2.1), we show that the Pitman estimator $\underline{\delta}_p$ is minimax. The proof makes use of an identity due to Stein (1973). The identity has been used by Stein and others for finding an unbiased estimator of the risk difference. We also introduce mixed estimators for a general k. However, the form of the estimator is complicated even for $k = 3$.

A result of Brown (1971) proves that $\underline{\delta}_p$ is inadmissible for $k \geq 3$. However, he does not give any dominating estimator. For $k = 3$ we propose some alternative estimators which may possibly ? improve $\underline{\delta}_p$. These estimators are obtained using Brown's (1979) technique and through empirical Bayes considerations. The risk functions of the proposed estimators are too complicated for making any theoretical comparisons. We also use the Brewster-Zidek (1974) technique to prove admissibility of $\underline{\delta}_p$ among its multiples and obtain a necessary and sufficient condition for its admissibility in another class of estimators.

In Section 4.3, we consider the estimation of the components of $\underline{\theta}$. When we have two normal populations with means θ_1 and θ_2, $\theta_1 \leq \theta_2$ and unequal known variances, Cohen and Sackrowitz (1970) proved that the component δ_{p2} of $\underline{\delta}_p$ for estimating θ_2 is minimax with respect to the squared error loss. However, the situation is different when we have three normal populations with means θ_1, θ_2 and θ_3, $\theta_1 \leq \theta_2 \leq \theta_3$ and equal known variances. We show that the components δ_{p1} and δ_{p3} of $\underline{\delta}_p$ for estimating θ_1 and θ_3 respectively are not minimax. We have not been able to prove or disprove the minimaxity of δ_{p2} as an estimator of θ_2.

Finally, we consider the estimation of the larger of two location parameters in the set up given below:

Let X_1 and X_2 be two independent random variables with the densities $f(x-\theta_1)$ and $f(x-\theta_2)$ respectively, $\theta_1 \leq \theta_2$ and also let f be symmetric about zero. Let $\delta(X_2) = X_2 + \alpha(X_2)$ be an estimator of θ_2 with $\alpha(X_2)$ bounded above. We prove the inadmissibility of $\delta(X_2)$ when the loss function is strictly convex. The result is a generalization of Theorems 2.1 and 5.1 of Cohen and Sackrowitz (1970

In Chapter 5, the general problem of estimation is considered. Let the underlying family of probability distributions be absolutely continuous with respect to a σ-finite measure μ and the loss function be quadratic. Suppose the estimation problem is invariant under a finite group G of transformations which preserve μ and the induced group $\tilde{G}$ be a commutative subgroup of the group of linear transformations. Moors (1981, 1985) obtained a sufficient condition for the inadmissibility of G-equivariant estimators. In a subsequent paper, Moors and Van Houwelingen (1987) showed that

the conditions that $g \in G$ preserve μ and that $\tilde{G}$ be commutative can be relaxed. In Section 5.2, we consider a more general set up of any locally compact group with $\tilde{G}$ a subgroup of the affine group and the loss function any strictly increasing function of the Euclidean distance between the estimate and the estimand. We obtain sufficient conditions for the inadmissibility of G-equivariant estimators. The conditions that $\tilde{G}$ be commutative and $g \in G$ be measurepreserving, shown to be redundant by Moors and Van Houwelingen (1987) for finite groups, seem to be necessary in the general set up.

In Section 5.3, some applications of the inadmissibility result are given. In particular, we consider the estimation of two or more ordered normal means and the estimation of two ordered binomial probabilities of success when the loss function is quadratic. We also consider the estimation of normal mean and binomial probability of success when these parameters are restricted to the bounded subsets of the natural parameter space and the loss function is quartic.

Chapter 6 deals with the problem of estimating the common location of two populations which have unequal and unknown scale parameters. An account of earlier work on this problem was given in Section 1.1.2. Let $X = (X_1,\ldots,X_m)$ and $\underline{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from populations with the densities $\frac{1}{\sigma_x} f(\frac{x-\theta}{\sigma_x})$ and $\frac{1}{\sigma_y} f(\frac{x-\theta}{\sigma_y})$ respectively, where σ_x and σ_y are unknown and possibly unequal. The density f is assumed to be an even function and the loss function squared error. Let Ψ_x and γ_x be odd location-scale and even location free scale estimators of θ and σ_x respectively based on $\underline{X}$. Similarly, let Ψ_y and γ_y be defined

for $\underline{Y}$. Cohen (1976) considered improving Ψ_x and suggested an unbiased estimator δ_a,

$$\delta_a = \left(1 - \frac{a}{1+Z}\right)\Psi_x + \frac{a}{1+Z}\Psi_y , \tag{1.2.5}$$

where $Z = \tau_y^2/\tau_x^2$ and a is a positive real. He obtained an upperbound $a^*(m,n)$ for a such that for $0 < a \leq a^*(m,n)$, δ_a dominates Ψ_x. Bhattacharyya (1981) enlarged the class of estimators δ_a dominating Ψ_x by finding a larger upperbound $A(m,n)$ on a, that is, $A(m,n)$ is such that for $0 < a \leq A(m,n)$, the estimator δ_a improves Ψ_x

We, in Section 6.3, develop sufficient conditions which provide bounds different from those of Cohen and Bhattacharyya. The bound in Theorem 6.3.2 is sharper than $A(m,n)$ and thus results in a larger class of estimators δ_a dominating Ψ_x. Also the bound is easier to calculate than $A(m,n)$. The bound in Theorem 6.3.1 is not always larger than $A(m,n)$, however, its calculation is much easier. We apply Theorems 6.3.1 and 6.3.2 to the uniform distribution and obtain bounds $B_1(m,n)$ and $B_4(m,n)$ respectively. It is seen that $B_1(m,n)$ is not always larger than $A(m,n)$ but $B_4(m,n)$ is never less than $A(m,n)$ for any values of m and n. In fact, $B_4(m,n) \geq 1$ for both pairs (m,n) and (n,m) when $m = n \geq 20$ and also when $n \geq 32$ if $m \leq n+5$ compared to $A(m,n) \geq 1$ for both pairs (m,n) and (n,m) when $m = n \geq 25$ and when $n \geq 35$ if $m \leq n+5$. As pointed out in Section 1.1.2, if the bound is greater than or equal to 1 for both pairs (m,n) and (n,m), the estimator δ_a, for $a = 1$, is better than both Ψ_x and Ψ_y.

In Section 6.4 a different combined estimator of θ is considered:

$$\delta_c^* = W_1 \Psi_x + W_2 \Psi_y , \tag{1.2.6}$$

where $W_1 = \dfrac{c\gamma_x^{-2}}{c\gamma_x^{-2} + \gamma_y^{-2}}$ and $W_2 = 1-W_1$. We develop sufficient conditions for δ_c^* to improve Ψ_x. Using a symmetry argument the classes of estimators dominating both Ψ_x and Ψ_y are obtained.

In Section 6.6, we consider improving the Graybill-Deal estimator $\hat{\mu}_{GD}$ for estimating the common mean μ of two normal distributions with unknown and unequal variances. Sinha and Mouqadem (1982) investigated admissibility of $\hat{\mu}_{GD}$ in certain subclasses of $\mathcal{C} = \{\hat{\mu}: \hat{\mu} = \bar{X} + D\,\Psi(S_1, S_2, D^2),\ 0 \le \Psi \le 1\}$, where $D = \bar{Y}-\bar{X}$. They also gave a simplified expression for the risk function of an estimator $\hat{\mu} \in \mathcal{C}$. We restrict attention to a subclass $\mathcal{C}_1 = \{\hat{\mu}: \hat{\mu} = \bar{X} + D\,\Psi(S_1, S_2),\ 0 \le \Psi \le 1\}$ of $\mathcal{C}$ and obtain a differential inequality. A solution of the differential inequality would provide an improvement over $\hat{\mu}_{GD}$ in $\mathcal{C}_1$. We have not been able to prove the existence of a solution and the question of inadmissibility of $\hat{\mu}_{GD}$ is still open.

In Section 6.7, the problem of estimating the common mean of a bivariate normal population with unknown correlation ρ is taken up. Krishnamoorthy and Rohatgi (1988a) proposed an estimator $\hat{\mu}_3$ and obtained the region of the parameter space where it dominates the MLE $\hat{\mu}_2$. They also compared the risk performances of $\hat{\mu}_2$ and $\hat{\mu}_3$. We consider some modifications of $\hat{\mu}_3$ which have larger region of dominance over $\hat{\mu}_2$ than $\hat{\mu}_3$. These modified estimators belong to a

larger class of estimators $\{\hat{\mu}_3(c),\ c \text{ real}\}$. We obtain minimal essentially complete class of estimators in the class. Finally the risk performances of some of these estimators are compared numerically with those of $\hat{\mu}_2$ and $\hat{\mu}_3$.

CHAPTER - 2

Basic Definitions and Results

In this chapter we give some basic definitions and results which will be used subsequently. For more details one can see Ferguson (1967), Lehmann (1983) and Berger (1985).

Let X be a random observable (usually a vector) with values in a space $\mathcal{X}$, and have a distribution P_θ, θ belonging to the parameter space Ω. The parameter θ is unknown but Ω is assumed to be known. The problem is of estimating a measurable function $h(\theta)$ of θ. We assume that the estimates lie in a space $\mathcal{A}$, which is the convex closure of $h(\Omega) = \{h(\theta): \theta \in \Omega\}$. $\mathcal{A}$ is also called the action space of the estimation problem.

Let $\delta(X)$ be an estimator of $h(\theta)$. When $X = x$ is observed, the loss incurred in estimating $h(\theta)$ by δ is $L(\theta, \delta(x))$, a real valued, nonnegative and measurable function in both its arguments. We assume $L(\theta, h(\theta)) = 0$, that is, the loss is zero whenever correct value is estimated.

The loss function $L(\theta, a)$ is said to be convex (strictly convex), if it is a convex (strictly convex) function of a. Usually, the loss function is taken to be an increasing function of the Euclidean distance $|h(\theta) - \delta(x)|$ between the estimate and the estimand. Two commonly used loss functions are

$$L_1(\theta, a) = |\theta - a|^2, \tag{2.1}$$

and $L'(\theta, a) = |\theta - a|$, (2.2)

called the squared error loss function and the absolute error loss function respectively.

The performance of an estimator $\delta(X)$ is measured by the risk function

$$R(\theta,\delta) = E_\theta L(\theta, \delta(X)),$$

where E_θ denotes the expectation when the true probability distribution is P_θ.

Ideally one would like to have a δ which has minimum value of $R(\theta,\delta)$ for all $\theta \in \Omega$. However, this is possible only when $h(\theta)$ is a constant or Ω is a singleton, both the cases being of no interest.

An estimator $\delta_o(X)$ is said to be better than $\delta_1(X)$ if

$$R(\theta,\delta_o) \leq R(\theta,\delta_1) \quad \text{for all } \theta \in \Omega ,$$

$$\text{and} \quad R(\theta',\delta_o) < R(\theta',\delta_1) \quad \text{for some } \theta' \in \Omega .$$

Equivalently, we say that δ_o improves δ_1 or δ_o dominates δ_1.

The estimator δ_o is said to be as good as δ_1 if

$$R(\theta,\delta_o) \leq R(\theta,\delta_1) \quad \text{for all } \theta \in \Omega.$$

Let the class of all the estimators be denoted by $\mathcal{D}$ and $\mathcal{C}$ be a subclass of $\mathcal{D}$. An estimator δ_o is said to be admissible in the class $\mathcal{C}$, if

$$R(\theta,\delta) \leq R(\theta,\delta_o) \quad \text{for all } \theta \in \Omega \text{ and any } \delta \in \mathcal{C}$$

$$\Rightarrow R(\theta,\delta) = R(\theta,\delta_o) \quad \text{for all } \theta \in \Omega.$$

Alternatively, δ_o is admissible in $\mathcal{C}$, if there is no estimator in $\mathcal{C}$ better than δ_o. An estimator admissible in $\mathcal{D}$ will simply be called admissible.

If an estimator is not admissible, it is said to be inadmissible. Thus, if δ_o is an inadmissible estimator, we can find a δ_1 better than δ_o.

A subclass $\mathcal{C}$ of $\mathcal{D}$ is said to be complete (essentially complete) if for any δ_1 not in $\mathcal{C}$, we can find a δ_2 in $\mathcal{C}$ which is better than (as good as) δ_1. We call $\mathcal{C}$ a minimal complete (minimal essentially complete) class, if $\mathcal{C}$ is complete (essentially complete) and no subclass of $\mathcal{C}$ is complete (essentially complete). If, in an estimation problem, a minimal complete class exists, it is the class of all the admissible estimators.

The advantage of having a complete (essentially complete) class is that we need not look outside it to find an estimator, for we have a better (as good as) estimator in it.

When the loss function is convex, the class D of all non-randomized estimators is essentially complete in $\mathcal{D}$. (For a definition of the randomized estimators, see Ferguson (1967).) Further, if a sufficient statistic T exists, then the class of all non-randomized estimators based on T is essentially complete.

An estimator δ_o is said to be minimax if

$$\operatorname*{Sup}_{\theta \in \Omega} R(\theta, \delta_o) = \operatorname*{Inf}_{\delta \in \mathcal{D}} \operatorname*{Sup}_{\theta \in \Omega} R(\theta, \delta) . \tag{2.3}$$

The value on the right hand side of (2.3) is called the minimax risk or the upper value of the estimation problem.

Let $\mathcal{B}(\Omega)$ be a σ-field of subsets of Ω. A measure π on $(\Omega, \mathcal{B}(\Omega))$ is said to be a proper prior if $\pi(\Omega) < \infty$.

If $\pi(\Omega) = \infty$ with

$$\int_{\Omega} p_\theta(x)\, d\pi(\theta) < \infty \quad \text{for almost all } x \text{ (a.e. } [\mu]) , \qquad (2.4)$$

where $p_\theta(x)$ is the density of $P_\theta(x)$ with respect to a σ-finite measure μ on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ and $\mathcal{B}(\mathcal{X})$ a σ-field of subsets of $\mathcal{X}$; π is called an improper prior on Ω.

The posterior or formal posterior distribution of θ given x is $p_\theta(x)\, d\pi(\theta) / \int_\Omega p_\theta(x)\, d\pi(\theta)$ according as the prior measure π is proper or improper. The corresponding posterior or formal posterior risk of δ is

$$\int_\Omega L(\theta, \delta(x))\, p_\theta(x)\, d\pi(\theta) \Big/ \int_\Omega p_\theta(x)\, d\pi(\theta) .$$

An estimator δ_π minimizing the posterior risk (formal posterior risk), or equivalently, $\int_\Omega L(\theta,\delta(x))\, p_\theta(x)\, d\pi(\theta)$ is called a Bayes (formal Bayes or generalized Bayes) estimator with respect to π. When $\pi(\Omega) < \infty$, δ_π also minimizes the Bayes risk defined by

$$r(\pi,\delta) = \int_\Omega R(\theta,\delta)\, d\pi(\theta) .$$

When the loss function is squared error, the Bayes estimator is the mean of the posterior distribution and when the loss is absolute error, the Bayes estimator is the median of the posterior distribution.

Let $\varepsilon > 0$. An estimator δ_o is said to be ε-Bayes with respect to π, if

$$r(\pi,\delta_o) \leq \inf_{\delta \in \mathcal{D}} r(\pi,\delta) + \varepsilon .$$

An estimator δ_o is extended Bayes, if for any $\varepsilon > 0$, there is a prior π_ε such that δ_o is ε-Bayes with respect to π_ε.

Let Ω^* be the class of all prior distributions on Ω. A prior distribution π^* is said to be least favourable, if

$$\inf_{\delta \in \mathcal{D}} r(\pi^*, \delta) = \sup_{\pi \in \Omega^*} \inf_{\delta \in \mathcal{D}} r(\pi,\delta) . \tag{2.5}$$

Thus, a least favourable prior maximizes the minimum Bayes risk. The expression on the right hand side of (2.5) is called the lower value of the estimation problem. It is always less than or equal to the upper value of (2.3). When the two are equal, the common value is called the (minimax) value of the problem. In Theorem 2.1 below, we state the conditions for an estimation problem to have a value. We need the notion of a lower semicontinuous function. A real valued function f is said to be lower semicontinuous, if for each $c \in R^1$, the set $\{x: f(x) > c\}$ is open.

Theorem 2.1 (Ferguson (1967, p. 85)): Let $\mathcal{C} \subset \mathcal{D}$ be an essentially complete class. Assume that there is a topology on $\mathcal{C}$ such that $\mathcal{C}$ is compact and $R(\theta,\delta)$ is lower semicontinuous in $\delta \in \mathcal{C}$ for all $\theta \in \Omega$. Then the estimation problem has a value and a minimax estimator exists.

In the problems with a value, we expect the Bayes estimator with respect to the least favourable prior to be minimax as well. This is seen to be true in general as stated below.

Theorem 2.2: Assume that the estimation problem has a value and that a minimax estimator exists as well as a least favourable prior π^*. If δ_o is the unique Bayes estimator with respect to π^*, then δ_o is minimax.

Another optimality criterion, sometimes used, is that of $\mathcal{J}$-minimaxity. Let $\mathcal{J}$ be a subclass of Ω^*. An estimator δ_o is said to be $\mathcal{J}$-minimax if

$$\sup_{\pi \in \mathcal{J}} r(\pi,\delta_o) = \inf_{\delta \in \mathcal{D}} \sup_{\pi \in \mathcal{J}} r(\pi,\delta) .$$

An $\mathcal{J}$-minimax estimator reduces to a Bayes estimator in case $\mathcal{J}$ is singleton and to a minimax estimator if $\mathcal{J}$ is the whole of Ω^*.

Next, we introduce invariance in estimation problems:

Let G denote a group of measurable transformations from $\mathcal{X}$ into itself. The group operation is composition: if g_1 and g_2 are transformations from $\mathcal{X}$ into itself, $g_2 g_1$ is defined as the transformation $x \to g_2(g_1(x))$.

The family $\mathcal{P} = \{P_\theta: \theta \in \Omega\}$ of probability distributions on $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$ is said to be invariant under G, if for every $g \in G$ and every $\theta \in \Omega$, there exists a unique θ' such that the distribution of g(X) is given by $P_{\theta'}$ whenever the distribution of X is given by P_θ. The θ' uniquely determined by g and θ is denoted by $\bar{g}(\theta)$. The set $\bar{G} = \{\bar{g}: g \in G\}$ is a group of transformations from Ω into itself. It is seen that $\bar{G}$ is a homomorphic image of G.

A loss function $L(\theta,a)$ is said to be invariant under G, if for every $g \in G$ and $a \in \mathcal{A}$ there exists an $a' \in \mathcal{A}$ such that

$$L(\theta,a) = L(\bar{g}(\theta),a') \quad \text{for all } \theta \in \Omega .$$

Without loss of generality a' can be chosen to be unique so that given a g it is a function of a. We denote a' by $\tilde{g}(a)$. The set $\tilde{G} = \{\tilde{g}: g \in G\}$ is again a group of transformations from $\mathcal{A}$ into itself and also it is homomorphic image of G and $\bar{G}$.

We call an estimation problem G-invariant, if the family of probability distributions $\mathcal{P}$ and the loss function $L(\theta, a)$ are invariant under G.

For a G-invariant problem, an estimator $\delta \in D$ is said to G-equivariant, if

$$\delta(g(x)) = \tilde{g}(\delta(x)) \quad \text{for all } x \in \mathfrak{X} \text{ and } g \in G .$$

Two points θ_1 and θ_2 of Ω are said to be equivalent, if there exists a $\bar{g} \in \bar{G}$ such that $\theta_1 = \bar{g}(\theta_2)$. This is an equivalence relation and the equivalence classes introduced by it are termed orbits of Ω under $\bar{G}$. A significant result is that a G-equivariant estimator has constant risk function on the orbits of Ω. Thus, in the problems where $\bar{G}$ is a transitive group, the risk of a G-equivariant estimator is independent of the parameter.

It is interesting to investigate the situations in which an estimator minimax among G-equivariant estimators is minimax among all the estimators. Kiefer (1957) obtained sufficient conditions for such a result to hold. The conditions imposed are both on the loss function and the group G of transformations. These conditions and the result are also discussed in Eaton (1966).

An application of Kiefer's theorem proves the minimaxity of the best translation equivariant estimator in estimation of a location vector.

<u>Theorem 2.3</u>: Let $\underline{X}_1,\ldots,\underline{X}_n$ be independent and identically distributed random variables with probability density $f(\underline{x}-\underline{\theta})$ on R^p, where $\theta \in R^p$. If $E|\underline{X}_i|^2 < \infty$, $i = 1,\ldots,n$; then the best translation equivariant estimator of $\underline{\theta}$ is minimax when the loss function is squared error.

In the set up of Theorem 2.3, consider estimation of a component θ_i of $\underline{\theta}$, $i = 1,\ldots,p$. It can be shown that for this problem also Kiefer's result is applicable and so we have the following result.

<u>Theorem 2.4</u>: In the set up of Theorem 2.3, the component of the best translation equivariant estimator for estimating θ_i; $i = 1,\ldots,p$, is minimax when the loss function is squared error.

Theorems 2.3 and 2.4 are used to prove general minimaxity results for the problem of estimating ordered location parameters considered in Chapter 3.

When the group G is compact, we can restrict attention to the class of G-equivariant estimators for proving admissibility or inadmissibility of G-equivariant estimators. Blackwell and Girshick (1954) and Savage (1954) originally proved this result for finite groups. The result in its generality, as stated below, was proved by Stein (1956).

<u>Theorem 2.5</u>: Let an estimation problem be invariant under a compact group G of transformations. Let δ_o be a G-equivariant admissible estimator and for any δ_1 satisfying

$$R(\theta,\delta_1) \leq R(\theta,\delta_o) \quad \text{for all } \theta \in \Omega\,,$$

$R(\bar{g}(\theta),\delta_1)$ be a continuous function of g for each θ. Then δ_o is an admissible estimator.

In Chapter 5, we consider G-equivariant estimation problems for specific distributions, when the group G is finite or compact. For each observed x, a subspace A_x of the action space $\mathcal{A}$ is obtained such that a G-equivariant estimator $\delta(X)$, not taking values in A_X

with positive probability for some $\theta \in \Omega$, is inadmissible. Thus, a G-equivariant admissible estimator, and so admissible by Theorem 2.5, must necessarily lie inside A_X with probability one.

Consider an estimation problem invariant under a group G of transformations. We are interested in finding an optimal G-equivariant estimator. Let T be a minimal sufficient statistic for Ω. It is a common practice to restrict attention to nonrandomized G-equivariant estimators based on T alone. Sharma (1983) proved under fairly general conditions that there is no loss of generality in this approach. Specifically, the question is the following: Is, D_T, the class of all nonrandomized G-equivariant estimators based on T, essentially complete in D_X, the class of all nonrandomized G-equivariant estimators? As an example, we consider the problem of estimating the common mean of two different normal populations:

Let $\underline{X} = (X_1,\ldots,X_m)$ and $\underline{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from $N(\mu,\sigma_1^2)$ and $N(\mu,\sigma_2^2)$ populations respectively. The variances σ_1^2 and σ_2^2 are assumed to be unequal and unknown. Consider estimation of μ when the loss is

$$L_2(\hat{\mu},\mu) = (\hat{\mu}-\mu)^2/\sigma_1^2 .$$

Let G be the affine group of transformations, that is, under G, $X_i \to aX_i+b$, $i = 1,\ldots,m$; $Y_j \to aY_j+b$, $j = 1,\ldots,n$; a and b are real and $a \neq 0$. Then $(\mu,\sigma_1^2,\sigma_2^2) \to (a\mu+b,\ a^2\sigma_1^2,\ a^2\sigma_2^2)$ and the estimation problem is invariant under G. The loss function is convex and the minimal sufficient statistic $T = (\bar{X},\ \bar{Y},\ S_1,\ S_2)$ is G-equivariant, that is,

$$T(X_1,\ldots,X_m,\ Y_1,\ldots,Y_n) = T(X_1',\ldots,X_m',\ Y_1',\ldots,Y_n')$$
$$\Rightarrow\ T(aX_1+b,\ldots,aX_m+b,\ aY_1+b,\ldots,aY_n+b)$$
$$= T(aX_1'+b,\ldots,aX_m'+b,\ aY_1'+b,\ldots,aY_n'+b)\ .$$

Now, the question is, can we, without any loss, base our nonrandomized G-equivariant estimators only on T? The result of Sharma (1983, Theorem 3.2) answers this question in the affirmative, since an affine transformation involves only arithmetic operations.

In Chapter 5, we use Haar measure to obtain an essentially complete class of G-equivariant estimators in the problem of estimating two normal means when there are no restrictions on the parameter space. We briefly describe below a Haar measure (for details see Halmos (1952)).

Consider a topological group G, that is, a group which is a Hausdorff space and the mapping of GxG into G, defined by $f(g_1,g_2) = g_1^{-1}g_2$, is continuous. Let G be locally compact and let $\mathcal{B}(G)$ be a class of subsets of G. A left invariant Haar measure is a σ-finite measure μ_1 on $(G, \mathcal{B}(G))$ satisfying

of G and unique upto a positive constant of proportionality (see Halmos (1952, p. 254)).

In Chapter 5, we consider Haar measure on the orthogonal group, that is, $X \to \Gamma X$, Γ an orthogonal matrix. A brief description of this is given below (see Lehmann (1959, p. 335)):

Consider the group G of kxk orthogonal matrices, with the matrix product $\Gamma_1\Gamma_2$ as the binary operation. G can be thought of as a subset of k^2-dimensional Euclidean space R^{k^2}, by representing a matrix as a point in R^{k^2}. Take a class of Borel subsets of G as the σ-field $\mathcal{B}(G)$. Now, G being a compact group, there exists a right invariant probability measure ν on $(G, \mathcal{B}(G))$ satisfying

$$\nu(B\Gamma) = \nu(B) \quad \text{for all } B \in \mathcal{B}(G) \text{ and } \Gamma \in G.$$

The measure ν can be constructed as follows:

Let $X = (X_{ij})$ be a kxk random matrix with X_{ij} independently and identically distributed N(0,1) random variables. The distributions of $X\Gamma'$ and X are the same for any orthogonal matrix Γ'. By an application of Gram-Schmidt orthogonalization process, we can, for each X, find an orthogonal matrix $Y = f(X)$ such that $Y\Gamma' = f(X\Gamma')$. Now the distribution of Y and $Y\Gamma'$ are the same and so the probability distribution ν of the random matrix Y is clearly orthogonal invariant.

When k = 2, the group G of orthogonal transformations can be described by

$$G = \left\{ \begin{pmatrix} \mathrm{Cos}\theta & \mathrm{Sin}\theta \\ \mp\mathrm{Sin}\theta & \pm\mathrm{Cos}\theta \end{pmatrix} : -\pi < \theta < \pi \right\}.$$

An orthogonal-invariant probability distribution on G is the uniform distribution over the interval $(-\pi, \pi)$.

Next, we describe some techniques for proving inadmissibility of estimators.

The Brewster-Zidek Technique

A useful technique for improving upon equivariant estimators was developed by Brewster and Zidek (1974). However, the method as such can be used in general. Suppose we want to improve an estimator δ^o. Consider a class $\mathcal{C}$ of estimators of which δ^o is a member. Let δ_θ be the optimal choice in $\mathcal{C}$ for each θ. A suitable choice of $\mathcal{C}$ may render δ_θ independent of θ. For example, in the problem of estimating the variance σ^2 of a normal population with unknown mean μ, based on a random sample $X_1,\ldots,X_n$, the maximum likelihood estimator S^2/n is improved by $S^2/(n+1)$, the best choice in the class $\{cS^2 \colon c > 0\}$.

Sometimes it is difficult to choose a $\mathcal{C}$ such that δ_θ is independent of θ. Improvements are still possible, if δ_θ does not vary significantly with θ, but is quite different from δ^o. Consider a class $\mathcal{F} = \{\delta_c \colon c \text{ real}\}$ of which δ^o is a member, say with $c = c_o$. Find $\hat{c}_\theta$, the choice of c minimizing $R(\theta,\delta_c)$ for each θ. If $R(\theta,\delta_c)$ is a strictly convex function of c, improvement over δ^o is possible provided $\underline{c} = \inf_{\theta \in \Omega} \hat{c}_\theta > c_o$ or $\bar{c} = \sup_{\theta \in \Omega} \hat{c}_\theta <$ The dominating estimator is $\delta_{\underline{c}}$ or $\delta_{\bar{c}}$ according as the first or second inequality holds. Further, the class $\{\delta_c \colon \underline{c} \le c \le \bar{c}\}$ is minimal complete in $\mathcal{F}$.

This method can be generalized to cases; where we consider classes of estimators characterized by two or more constants. As an example, consider estimators of the form $\delta_{c,d}$, where $(c,d) = (c_o,d_o)$ gives δ^o. Assume that the risk function $R(\theta,\delta_{c,d})$ is

strictly bowl-shaped (see Brewster and Zidek (1974) for a definition) and that $\hat{c}_\theta$ and $\hat{d}_\theta$ minimizing $R(\theta, \delta_{c,d})$ for each θ can be obtained explicitly. Call $\underline{c} = \operatorname*{Inf}_{\theta \in \Omega} \hat{c}_\theta$, $\bar{c} = \operatorname*{Sup}_{\theta \in \Omega} \hat{c}_\theta$, $\underline{d} = \operatorname*{Inf}_{\theta \in \Omega} \hat{d}_\theta$ and $\bar{d} = \operatorname*{Sup}_{\theta \in \Omega} \hat{d}_\theta$. Then an estimator dominating δ^o can be obtained if $\underline{c} > c_o$, $\underline{d} > d_o$ or $\bar{c} < c_o$ and $\bar{d} < d_o$.

The orbit by orbit improvement technique of Brewster and Zidek involves reducing the risk of an equivariant estimator on the orbits of some invariant statistic Z, usually taken to be a maximal invariant. Consider estimators of the form $\delta_{\Psi(Z)}$, where $\Psi = \Psi^o$ gives δ^o. Define the conditional risk function of $\delta_{\Psi(Z)}$ given $Z = z$ as

$$R(\theta, \delta_{\Psi(z)}) = E\{L(\theta, \delta_{\Psi(Z)}(X)) \mid Z = z\} .$$

Assume that $R(\theta, \delta_{\Psi(z)})$ is a strictly convex function of $\Psi(z)$ and that $\hat{\Psi}_\theta(z)$ minimizes $R(\theta, \delta_{\Psi(z)})$ for each θ and z. Identify the sets

$$A = \{z: \underline{\Psi}(z) = \operatorname*{Inf}_{\theta \in \Omega} \hat{\Psi}_\theta(z) > \Psi_o(z)\}$$

and $$B = \{z: \bar{\Psi}(z) = \operatorname*{Sup}_{\theta \in \Omega} \hat{\Psi}_\theta(z) < \Psi_o(z)\} .$$

Then the estimator $\delta_{\Psi^*(Z)}$, where

$$\begin{aligned} \Psi^*(z) &= \underline{\Psi}(z) , \quad \text{if } z \in A \\ &= \bar{\Psi}(z) , \quad \text{if } z \in B \\ &= \Psi^o(z) \text{ elsewhere,} \end{aligned}$$

dominates δ^o provided $P_\theta\,(A \cup B) > 0$ for some $\theta \in \Omega$.

<u>Brown's Heuristic Approach</u>

Brown (1979) developed a heuristic method for proving admissibility or inadmissibility of estimators. He reduced the

question of admissibility to the problem of finding solutions of certain differential inequalities. The coefficients of these inequalities involve moments of the underlying distributions.

Let $(\underline{X},\underline{Y})$ be a random vector with density $p_{\underline{\theta},\underline{\eta}}(\underline{x},\underline{y})$ with respect to Lebesgue measure on $R^k x R^l$. The parameters $\underline{\theta}$ and $\underline{\eta}$, are unknown with $(\underline{\theta},\underline{\eta}) \in \Omega \subset R^k x R^l$, $k \geq 1$, $l \geq 0$. The problem is to estimate $\underline{\theta}$ with a loss $\sum_{i=1}^{k} W(\theta_i - a_i)$, a sum of individual losses in estimating θ_i. Only nonrandomized estimators will be considered. Brown derived the inequalities for $k = 1$, but they have easy generalizations to higher dimensional cases. Below we give an inequality for $k = 3$ and $l = 0$ as used in Chapter 4.

Let $\underline{X} = (X_1,X_2,X_3)$ and the problem be of investigating admissibility of an estimator $\underline{\delta}(\underline{X}) = (\delta_1(\underline{X}),\delta_2(\underline{X}),\delta_3(\underline{X}))$. Define $\underline{\gamma}(\underline{X}) = \underline{\delta}(\underline{X}) - \underline{X}$ and consider an estimator $\underline{\delta}^*(\underline{X}) = \underline{\delta}(\underline{X}) + \underline{\lambda}(\underline{X})$, $\underline{\lambda}(\underline{X}) = (\lambda_1(\underline{X}),\lambda_2(\underline{X}),\lambda_3(\underline{X}))$. The risk difference between $\underline{\delta}$ and $\underline{\delta}^*$ is

$$\Delta(\underline{\theta},\underline{\eta}) = R(\underline{\theta},\underline{\eta},\underline{\delta}) - R(\underline{\theta},\underline{\eta},\underline{\delta}^*)$$

$$= \sum_{i=1}^{3} E_{(\underline{\theta},\underline{\eta})} \left[W(X_i+\gamma_i(\underline{X})-\theta_i) - W(X_i+\gamma_i(\underline{X}) + \lambda_i(\underline{X}) - \theta_i)\right]$$

In the derivation below W', W" and W''' denote the first, the second and the third order derivatives of W respectively. Also $\lambda_{ij}(\underline{x})$ denotes the derivative of $\lambda_i(\underline{x})$ with respect to x_j, $i,j = 1,2,3$.

A Taylor expansion of $W(X_i + \gamma_i(\underline{X}) + \lambda_i(\underline{X}) - \theta_i)$ about $X_i+\gamma_i(\underline{X})-\theta_i$ after ignoring the terms of third and higher order derivatives of W, $i = 1,2,3$, gives

$$\Delta(\underline{\theta},\underline{\eta}) = \sum_{i=1}^{3} E_{(\underline{\theta},\underline{\eta})} \left[-\lambda_i(\underline{X})W'(X_i+\gamma_i(\underline{X})-\theta_i) - \tfrac{1}{2}\lambda_i^2(\underline{X})W''(X_i+\gamma_i(\underline{X})-\theta_i)\right].$$

Also

$$W'(X_i+\gamma_i(\underline{X})-\theta_i) = W'(X_i-\theta_i) + \gamma_i(\underline{X})\ W''(Z_i))$$

and $$W''(X_i+\gamma_i(\underline{X})-\theta_i) = W''(X_i+\gamma_i(\underline{X})-\theta_i) + \gamma_i(\underline{X})\ W'''(Z_i') ,$$

where Z_i and Z_i' are points lying between $X_i-\theta_i$ and $X_i+\gamma_i(\underline{X})-\theta_i$. Ignoring the term involving W''',

$$\Delta(\underline{\theta},\underline{\eta}) = \sum_{i=1}^{3} E_{(\underline{\theta},\underline{\eta})}\ [-\lambda_i(\underline{X})\ \{W'(X_i-\theta_i) + \gamma_i(\underline{X})W''(Z_i)\}$$
$$- \tfrac{1}{2}\lambda_i^2(\underline{X})\ W''(X_i-\theta_i)] .$$

Notice that for squared error loss the above expression of Δ is exact. Further, expanding $\lambda_i(\underline{X})$ and $\lambda_i^2(\underline{X})$ about $\underline{\theta}$ in Taylor series,

$$\lambda_i(\underline{X}) = \lambda_i(\underline{\theta}) + \sum_{j=1}^{3} (X_j-\theta_j)\ \lambda_{ij}(\underline{\theta}) + \ldots$$

and $$\lambda_i^2(\underline{X}) = \lambda_i^2(\underline{\theta}) + \ldots .$$

This yields

$$\Delta(\underline{\theta},\underline{\eta}) = \sum_{i=1}^{3} E_{(\underline{\theta},\underline{\eta})}\ [-\lambda_i(\underline{\theta})\ W'(X_i-\theta_i) - \sum_{j=1}^{3} (X_j-\theta_j)\lambda_{ij}(\underline{\theta})W'(X_i-\theta_i)$$

$$- \lambda_i(\underline{\theta})\gamma_i(\underline{X})W''(Z_i) - \sum_{j=1}^{3} (X_j-\theta_j)\lambda_{ij}(\underline{\theta})\gamma_i(\underline{X})W''(Z_i)$$

$$- \tfrac{1}{2}\lambda_i^2(\underline{X})W''(X_i-\theta_i) + \ldots]$$

Finally, $\gamma_i(\underline{X}) = \gamma_i(\underline{\theta}) + \ldots$ and as an approximation we take

$$\Delta(\underline{\theta},\underline{\eta}) = \sum_{i=1}^{3} E_{(\underline{\theta},\underline{\eta})}\ [-\lambda_i(\underline{\theta})W'(X_i-\theta_i) - \sum_{j=1}^{3}(X_j-\theta_j)\lambda_{ij}(\underline{\theta})W'(X_i-\theta_i)$$

$$- \lambda_i(\underline{\theta})\gamma_i(\underline{\theta})W''(Z_i) - \sum_{j=1}^{3}(X_j-\theta_j)\lambda_{ij}(\underline{\theta})\gamma_i(\underline{\theta})W''(Z_i)$$

$$- \frac{1}{2}\ \lambda_i^2(\underline{\theta})W''(X_i-\theta_i)] . \tag{2.6}$$

If there is a $\underline{\lambda}$ for which $\Delta(\underline{\theta},\underline{\eta})$ of (2.6) is positive for all $(\underline{\theta},\underline{\eta}) \in \Omega$ and the error in the above approximation is negligible, we can expect $\underline{\delta}^*$ to be better than $\underline{\delta}$. This is what we have attempted in Chapter 4 for obtaining an improvement over the Pitman estimator $\underline{\delta}_p$ of $\underline{\theta} = (\theta_1,\theta_2,\theta_3)$, where θ_i's are the means of independent normal populations with common variance unity and $\theta_1 \leq \theta_2 \leq \theta_3$.

Finally, we mention a result of Caridi(1983) about the completeness of the class of all Bayes and generalized Bayes estimators. Sacks (1963) was the first to prove the completeness for estimation in one parameter exponential family with a general loss function satisfying certain regularity conditions. Brown (1971) established this result for a p-variate normal distribution with the identity as covariance matrix when the loss function is quadratic while Berger and Srinivasan (1978) considered a p-variate exponential family and squared error loss. Caridi generalized this result to a p-variate exponential family with a general loss function. His set up is briefly described below:

Let $\underline{X}$ be p-dimensional random vector with density $p_{\underline{\theta}}(\underline{x})$ with respect to a σ-finite measure μ on R^p, given by $p_{\underline{\theta}}(\underline{x}) = c(\underline{\theta})e^{(\underline{x},\underline{\theta})}$, $(.,.)$ denoting the usual inner product. The parameter space Ω is any subset of the natural parameter space

$$\{\underline{\theta} \in R^p : \int_{R^p} e^{(\underline{x},\underline{\theta})} d\mu(\underline{x}) < \infty\}.$$

Under certain regularity conditions on the loss function and the measure μ, Caridi proved that if Ω is closed, the class of all Bayes and generalized Bayes estimators is complete.

In Chapter 3, Caridi's result is used to prove inadmissibility of the MLE and the mixed estimators of two ordered normal means.

Throughout the thesis φ and Φ stand for the standard normal probability density and standard normal probability distribution function respectively. Also $N_m(\underline{\gamma}, \Sigma)$ denotes the m-dimensional normal distribution with mean vector $\underline{\gamma}$ and the covariance matrix Σ. However, a univariate normal distribution with mean μ and variance σ^2 will be simply denoted by $N(\mu,\sigma^2)$.

CHAPTER - 3

Simultaneous Estimation of Ordered Location Parameters

3.1 Introduction

Let $(X_{i1},\ldots,X_{in})$ be a random sample from a continuous population with the density $f_i(x-\theta_i)$, $i = 1,\ldots,k$. Assume that the k samples are drawn independently and that θ_i's are ordered, say $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$. We take $\int_{-\infty}^{\infty} x\, f_i(x)dx = 0$, $i = 1,\ldots,k$, without loss of generality. For, if $\int_{-\infty}^{\infty} x\, f_i(x)dx = \mu_i$, $i = 1,\ldots,k$, define $X_{ij}' = X_{ij}-\mu_i$, $j = 1,\ldots,n$, $i = 1,\ldots,k$. The density of X_{ij}' is $f_i(x-\theta_i+\mu_i) = g_i(x-\theta_i)$, say, so that $\int_{-\infty}^{\infty} x\, g_i(x)dx = 0$, $i = 1,\ldots,k$. The problem is of estimating $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ when the loss function is squared error, that is,

$$L_1(\underline{\theta},\underline{a}) = |\underline{\theta}-\underline{a}|^2 = \sum_{i=1}^{k} (\theta_i-a_i)^2 . \tag{3.1.1}$$

When there are no restrictions on the parameter space, an estimator frequently used is the best translation equivariant estimator $\underline{X} = (X_1,\ldots,X_k)$, also called the Pitman estimator, where

$$X_i = \frac{\int_{-\infty}^{\infty} \theta_i \prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_i}{\int_{-\infty}^{\infty} \prod_{j=1}^{n} f_i(X_{ij}-\theta_i)d\theta_i}, \quad i = 1,\ldots,k . \tag{3.1.2}$$

Note that $\underline{X}$ is also the generalized Bayes estimator of $\underline{\theta}$ with respect to the uniform prior on R^k.

When the parameter space is restricted, we can use the analogue of the Pitman estimator, which is the generalized Bayes

estimator of $\underline{\theta}$ with respect to the uniform prior on the restricted space. When θ_i's are ordered we denote this estimator by $\underline{\delta}_p = (\delta_{p1},\ldots,\delta_{pk})$, where

$$\delta_{pr}(X_{11},\ldots,X_{kn}) = \frac{\int\limits_{\theta_1\le\ldots\le\theta_k}\cdots\int \theta_r \prod_{i=1}^{k}\prod_{j=1}^{n} f_i(X_{ij}-\theta_i)\, d\theta_1,\ldots,d\theta_k}{\int\limits_{\theta_1\le\ldots\le\theta_k}\cdots\int \prod_{i=1}^{k}\prod_{j=1}^{n} f_i(X_{ij}-\theta_i)\, d\theta_1,\ldots,d\theta_k}$$

$$r = 1,\ldots,k . \qquad (3.1.3)$$

Blumenthal and Cohen (1968b), for $k = 2$ and identical f_i's, proved that any estimator with the risk function never exceeding the constant risk of $\underline{X}$ is minimax. They use the fact that $\underline{X}$ is minimax for the unrestricted parameter space. In Section 3.2, their result is generalized to an arbitrary k and f_i's not necessarily the same. We also prove that for estimating a component θ_i of $\underline{\theta}$, any estimator with the risk function always below that of X_i is minimax.

In Section 3.3, we consider two normal populations with means θ_1 and θ_2, $\theta_1 \le \theta_2$, and the common variance unity. For estimating $\underline{\theta} = (\theta_1,\theta_2)$, we give a class of minimax estimators, all of which are improvements over the usual estimator $\underline{X}$. We also obtain a minimal essentially complete class of estimators among these estimators. However, these estimators are inadmissible. We do not know what estimators improve upon them. Another class of minimax estimators is also given. An estimator in the former class turns out to be the best choice in the latter. A comparison of this estimator with the Pitman estimator $\underline{\delta}_p$ is made.

In Section 3.4, we take up two normal populations with unequal but known variances and obtain two classes of minimax

estimators. The maximum likelihood estimator is shown to improve some of the minimax estimators. Also, all of these estimators, like the ones in Section 3.3, are shown to be inadmissible. We also prove the minimaxity of the Pitman estimator $\underline{\delta}_p$.

Blumenthal and Cohen (1968b) obtained sufficient conditions for the admissibility and minimaxity of the Pitman estimator $\underline{\delta}_p$ when we have two populations with densities $f(x-\theta_1)$ and $f(x-\theta_2)$. In Section 3.5, we consider densities $f_1(x-\theta_1)$ and $f_2(x-\theta_2)$ for the two populations, f_1 and f_2 not necessarily the same and develop sufficient conditions for the minimaxity of $\underline{\delta}_p$. Some applications of these results are also given.

3.2 Generalization of a Theorem of Blumenthal and Cohen

We first introduce notation for the general problem considered in this section.

Let $X_{i1},\ldots,X_{in}$, $i = 1,\ldots,k$ be real valued and independent random variables with X_{ij} having the density $f_i(x-\theta_i)$, $j = 1,\ldots,n$ with respect to Lebesgue measure. Assume

$$\int_{-\infty}^{\infty} x\, f_i(x)\, dx = 0\ , \qquad i = 1,\ldots,k,$$

without loss of generality. We want to estimate $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ under the condition $\theta_1 \le \theta_2 \le \cdots \le \theta_k$. Denote by Ω_o the restricted parameter space $\{(\theta_1,\ldots,\theta_k): \theta_1 \le \theta_2 \le \cdots \le \theta_k\}$. Let the loss function $L(\underline{\theta},\underline{a})$ and the estimators $\underline{X}$ and $\underline{\delta}_p$ be as in (3.1.1), (3.1.2) and (3.1.3) respectively. Also define

$$\underline{Y}_i = (Y_{i1},\ldots,Y_{i,n-1}), \quad \text{where } Y_{ij} = X_{i,j+1}-X_{i1};\ j = 1,\ldots,n,$$
$$i = 1,\ldots,k.$$

X_i can also be written as

$$X_i = X_{i1} - r_{1i}(\underline{Y}_i), \tag{3.2.1}$$

where

$$r_{1i}(\underline{y}_i) = \frac{\int_{-\infty}^{\infty} x\, f_i(x)\, f_i(x+y_{i1}) \ldots f_i(x+y_{i,n-1})dx}{r_{oi}(\underline{y}_i)}, \tag{3.2.2}$$

and

$$r_{oi}(\underline{y}_i) = \int_{-\infty}^{\infty} f_i(x)\, f_i(x+y_{i1}) \ldots f_i(x+y_{i,n-1})dx, \quad i = 1,\ldots,k. \tag{3.2.3}$$

Then the conditional probability density of X_i given Y_i is $p_i(x_i-\theta_i, y_i)$, where

$$p_i(x_i, \underline{y}_i) = \frac{f_i(x_i+r_{1i}(\underline{y}_i))\, f_i(x_i+r_{1i}(\underline{y}_i)+y_{i1}) \ldots f_i(x_i+r_{1i}(\underline{y}_i)+y_{i,n-1})}{r_{oi}(\underline{y}_i)} \tag{3.2.4}$$

Let $\underline{Y}$ denote the $k(n-1)$-vector $(\underline{Y}_1,\ldots,\underline{Y}_k)$. Since we have a one-to-one transformation from $(X_{i1},\ldots,X_{in})$ to $(X_i, Y_{i1},\ldots,Y_{i,n-1}) \equiv (X_i, \underline{Y}_i)$, $i = 1,\ldots,k$, we can base estimators of $\underline{\theta}$ on $(\underline{X}, \underline{Y})$.

Hereafter, $\mathcal{Y}_i$, $i = 1,\ldots,k$ will be arbitrary space, $\nu_i(.)$ a probability measure defined on the Borel subsets of $\mathcal{Y}_i$, $p_i(.,.) \geq 0$ on $R^1 \times \mathcal{Y}_i$ to R^1 jointly measurable in the two variables and

$$\int_{-\infty}^{\infty} p_i(x, \underline{y}_i)dx = 1 \quad \text{for all } \underline{y}_i \in \mathcal{Y}_i,\ i = 1,\ldots,k, \tag{3.2.5}$$

$$\int_{-\infty}^{\infty} xp_i(x, \underline{y}_i)dx = 0 \quad \text{for all } \underline{y}_i \in \mathcal{Y}_i,\ i = 1,\ldots,k. \tag{3.2.6}$$

It can be verified easily that the density given by (3.2.4) satisfies (3.2.5) and (3.2.6) with $\mathcal{Y}_i = R^{n-1}$ and $\nu_i(d\underline{y}_i)$ given by $r_{oi}(\underline{y}_i)dy_{i1}\ldots dy_{i,n-1}$.

We will write $\mathcal{Y}$ for $\mathcal{Y}_1 \times \ldots \times \mathcal{Y}_k$, $\underline{y}$ for $(\underline{y}_1,\ldots,\underline{y}_k)$, ν for $\nu_1 \times \ldots \times \nu_k$.

From now on the observed variables are $(X_i,\underline{Y}_i)$ with $X_i \in R^1$, $\underline{Y}_i \in \mathcal{Y}_i$, $i = 1,\ldots,k$ and the conditional density of X_i given $\underline{Y}_i$ is $p_i(x_i-\theta_i,\underline{y}_i)$, $\theta_i \in R^1$, $i = 1,\ldots,k$.

In terms of the new variables the risk of an estimator $\underline{\delta}(\underline{X},\underline{Y}) = (\delta_1(\underline{X},\underline{Y}),\ldots,\delta_k(\underline{X},\underline{Y}))$ for $\underline{\theta} = (\theta_1,\ldots,\theta_k)$ is

$$R(\underline{\theta},\underline{\delta}) = \int \ldots \int \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \sum_{i=1}^{k} (\delta_i(\underline{x},\underline{y})-\theta_i)^2 \prod_{i=1}^{k} p_i(x_i-\theta_i,\underline{y}_i)dx_1\ldots dx_k \, \nu_1(d\underline{y}_1)\ldots \nu_k(d\underline{y}_k). \tag{3.2.7}$$

Also, the Pitman estimator $\underline{\delta}_p = (\delta_{p1},\ldots,\delta_{pk})$ of (3.1.3) is given by

$$\delta_{pr}(\underline{X},\underline{Y}) = \frac{\int\limits_{\theta_1\leq\ldots\leq\theta_k} \ldots \int \theta_r \prod_{i=1}^{k} p_i(X_i-\theta_i,\underline{Y}_i)\, d\theta_1\ldots d\theta_k}{\int\limits_{\theta_1\leq\ldots\leq\theta_k} \ldots \int \prod_{i=1}^{k} p_i(X_i-\theta_i,\underline{Y}_i)\, d\theta_1\ldots d\theta_k}, \quad r = 1,\ldots,k. \tag{3.2.8}$$

Now we are ready to state the main theorem.

Theorem 3.2.1: Let Σ_o be a subset of R^k such that there exists a sequence $\{(a_{n1},\ldots,a_{nk}): n \geq 1\}$ for which

$$\liminf_{n \to \infty} \{(\theta_1,\ldots,\theta_k): (\theta_1+a_{n1},\ldots,\theta_k+a_{nk}) \in \Sigma_o\} = R^k . \tag{3.2.9}$$

If $\underline{\delta}$ is an estimator with

$$R(\underline{\theta},\underline{\delta}) \leq R < \infty \quad \text{whenever } \underline{\theta} \in \Sigma_o , \tag{3.2.10}$$

where R is the constant risk of the estimator $\underline{X}$, then $\underline{\delta}$ is a minimax estimator of $\underline{\theta}$ for $\underline{\theta} \in \Sigma_o$.

Proof: The proof of the theorem is essentially the same as that of Theorem 3.1 of Blumenthal and Cohen (1968b). To take care of different densities f_i, $i = 1,\ldots,k$, however, some necessary changes are made.

Let $\underline{\delta}$ be not minimax. Then there exists an estimator $\underline{\delta}^o = (\delta_1^o,\ldots,\delta_k^o)$ such that

$$\sup_{\underline{\theta} \in \Sigma_o} R(\underline{\theta},\underline{\delta}^o) < \sup_{\underline{\theta} \in \Sigma_o} R(\underline{\theta},\underline{\delta}) \leq R .$$

Therefore there exists an $\varepsilon > 0$ such that

$$R(\underline{\theta},\underline{\delta}^o) \leq R - \varepsilon \qquad \text{for all } \underline{\theta} \in \Sigma_o . \tag{3.2.11}$$

Let $\underline{\delta}'(\underline{a}_n) = (\delta_1'(\underline{a}_n),\ldots,\delta_k'(\underline{a}_n))$ be an estimator defined by

$$\delta_i'(\underline{a}_n,\underline{x},\underline{y}) = \delta_i^o(x_1+a_{n1},\ldots,x_k+a_{nk},\underline{y}) - a_{ni}, \quad i = 1,\ldots,k$$

and $(\theta_1+a_{n1},\ldots,\theta_k+a_{nk})$ belong to Σ_o. The inequality (3.2.11) and a change of variables in the integrand of (3.2.7), then, gives

$$R(\underline{\theta},\underline{\delta}'(\underline{a}_n)) \leq R - \varepsilon . \tag{3.2.12}$$

Following remarks of LeCam (1955, p. 80) one can show that when the parameter space is R^k, the class of all the estimators is compact after a compactification of the space $\mathcal{A}$ of estimates. Using this fact, condition (3.2.9) and inequality (3.2.12), we can, then, choose a subsequence of $\{\underline{\delta}'(\underline{a}_n)\}$ which converges weakly in limit to an estimator $\underline{\delta}^* = (\delta_1^*,\ldots,\delta_k^*)$ satisfying

$$R(\underline{\theta},\underline{\delta}^*) \leq R - \varepsilon \qquad \text{for all } \underline{\theta} \in R^k . \tag{3.2.13}$$

Now, for the unrestricted problem $\underline{X}$ is a minimax estimator (see Theorem 2.3, Chapter 2) and so from (3.2.13) we conclude that $\varepsilon = 0$. This proves the minimaxity of $\underline{\delta}$.

<u>Corollary 3.2.2</u>: Let $\underline{\delta}$ be an estimator of $\underline{\theta}$ with

$$R(\underline{\theta}, \underline{\delta}) \leq R \quad \text{for } \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k,$$

where R is the same as in Theorem 3.2.1. Then $\underline{\delta}$ is minimax for $\underline{\theta}$ when $\theta_1 \leq \cdots \leq \theta_k$.

<u>Proof</u>: Take $\Sigma_o = \Omega_o$ and $a_{ni} = (i-1)n$, $i = 1,\ldots,k$ in Theorem 3.2.1.

<u>Remark 3.2.1</u>: Taking $\underline{\delta} = \underline{X}$ in Theorem 3.2.1, we see that $\underline{X}$ is also a minimax estimator of $\underline{\theta}$ for $\underline{\theta} \in \Sigma_o$.

In the preceding set up, let us now consider estimation of a component of $\underline{\theta}$, say θ_i, with loss function the squared error. Then X_i is minimax when the parameter space is unrestricted (see Theorem 2.4, Chapter 2). Proceeding as in the proof of Theorem 3.2.1 we can prove the following result.

<u>Theorem 3.2.3</u>: Let $\delta_i(\underline{X}, \underline{Y})$ be an estimator of θ_i with

$$R(\underline{\theta}, \delta_i) \leq R_i < \infty \quad \text{for } \theta_1 \leq \theta_2 \leq \cdots \leq \theta_k,$$

where R_i is the constant risk of the estimator X_i. Then δ_i is minimax for θ_i when $\theta_1 \leq \cdots \leq \theta_k$.

3.3 <u>Two Normal Populations with Equal Variances</u>

In this section, we give two classes of minimax estimators, all of them improving upon the usual estimator. An essentially complete class of estimators among these estimators is also obtained.

These estimators are also proved to be not generalized Bayes and so according to a result of Caridi (1983) are inadmissible.

The Mixed Estimators

We start with a lemma due to Katz (1963, Theorem 1).

Lemma 3.3.1: Let $\underline{\delta} = (\delta_1, \delta_2)$ be an estimator of $\underline{\theta} = (\theta_1, \theta_2)$ when the parameter space is $\Omega_o = \{\underline{\theta}: \theta_1 \leq \theta_2\}$. Let the loss function be of the form

$$W(\delta_1 - \theta_1) + W(\delta_2 - \theta_2) , \tag{3.3.1}$$

where W is a convex, even and nonnegative function. Then if $P_{\underline{\theta}'}(\delta_1 > \delta_2) > 0$ for some $\underline{\theta}' \in \Omega_o$, $\underline{\delta}$ is dominated by

$$\underline{\delta}_{\alpha^+} = (\alpha\delta_1 + (1-\alpha)\delta_2,\ (1-\alpha)\delta_1 + \alpha\delta_2),$$

$$\text{where} \quad \begin{aligned} \alpha &= 1, && \text{if } \delta_1 \leq \delta_2 \\ &= \alpha^+, && \text{if } \delta_1 > \delta_2, \end{aligned} \tag{3.3.2}$$

α^+ is a constant lying between zero and one.

Remark 3.3.1: The estimator $\underline{\delta}_{\alpha^+}$ of (3.3.2) is called the mixed estimator of $\underline{\delta}$. Hereafter we will use the term 'mixed estimator' for the mixed estimator of the usual estimator $\underline{X}$.

Remark 3.3.2: The proof of the lemma follows from the fact that

$$W(\alpha^+\delta_1 + (1-\alpha^+)\delta_2 - \theta_1) + W((1-\alpha^+)\delta_1 + \alpha^+\delta_2 - \theta_2)$$
$$< W(\delta_1 - \theta_1) + W(\delta_2 - \theta_2)$$

whenever $\delta_1 > \delta_2$, $\theta_1 \leq \theta_2$ and $0 \leq \alpha^+ < 1$.

Remark 3.3.3: The estimator $\underline{\delta}_{1/2}$ which improves $\underline{\delta}$ is the projection of $\underline{\delta}$ on the space Ω_o. A more general result of this type is stated below.

Lemma 3.3.2: Let $h(\theta)$, the estimand, belong to R^m, the loss function be a strictly increasing function of the Euclidean distance $|h(\theta)-\delta(x)|$ between $h(\theta)$ and $\delta(x)$ and the estimation problem (Ω,L,X) be restricted to (Ω^*,L,X), $\Omega^* \subset \Omega$. Denote the convex closure of $h(\Omega^*)$ by $\mathcal{A}_o$. Then any estimator $\delta(X)$, satisfying $P_{\theta'}(\delta(X) \notin \mathcal{A}_o) > 0$ for some $\theta' \in \Omega^*$ is inadmissible.

Proof: We define
$$\delta'(x) = \delta(x), \text{ if } \delta(x) \in \mathcal{A}_o$$
$$= \delta_o(x), \text{ if } \delta(x) \notin \mathcal{A}_o,$$

where $\delta_o(x)$ is the perpendicular projection of $\delta(x)$ on $\mathcal{A}_o$.

Now, whenever $\delta(x) \notin \mathcal{A}_o$, we have

$$|\delta_o(x)-h(\theta')| < |\delta(x)-h(\theta')|$$

by definition of a perpendicular projection (see Lemma 5.2.5). Hence $R(\theta',\delta)-R(\theta',\delta')$

$$= \int [L(|h(\theta')-\delta(x)|) - L(|h(\theta')-\delta'(x)|)]\, dP$$

$$= \int_{\delta(x) \in \mathcal{A}_o} + \int_{\delta(x) \notin \mathcal{A}_o} [L(|h(\theta')-\delta(x)|)-L(|h(\theta')-\delta'(x)|)]\, dP$$

> 0. This completes the proof of the lemma.

Let X_1 and X_2 be independent normal random variables with means θ_1 and θ_2 respectively, $\theta_1 \leq \theta_2$ and common variance unity. The mixed estimator $\underline{\delta}_{\alpha^+}(\underline{X})$ for $\underline{\theta} = (\theta_1,\theta_2)$ is given by

$$\underline{\delta}_{\alpha^+}(\underline{X}) = (\alpha X_1 + (1-\alpha)X_2,\ (1-\alpha)X_1 + \alpha X_2), \tag{3.3.3}$$

where $\quad \alpha = 1 \quad$ when $X_1 \leq X_2$

$\qquad\quad = \alpha^+ \quad$ when $X_1 > X_2$, α^+ real.

We show below that $\underline{\delta}_{\alpha^+}(\underline{X})$ for $0 \leq \alpha^+ < 1$ dominates $\underline{X} = (X_1, X_2)$.

The risk function of $\underline{\delta}_{\alpha^+}(\underline{X})$ is

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) = E(\alpha X_1 + (1-\alpha)X_2 - \theta_1)^2 + E((1-\alpha)X_1 + \alpha X_2 - \theta_2)^2$$

$$= 2 + 2E(1-\alpha)^2 Z^2 - 2E(1-\alpha)\ Z(Z-\eta),$$

where $\quad Z = X_1 - X_2 \sim N(\eta, 2)$,

$$\eta = \theta_1 - \theta_2 \leq 0 \quad \text{for } \theta_1 \leq \theta_2. \tag{3.3.4}$$

Some further simplification yields

$-ve \qquad -ve \quad \eta \leq 0$

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) = 2 - 2\alpha^+(1-\alpha^+) \int_{Z>0} Z^2 dP + 2\eta(1-\alpha^+) \int_{Z>0} Z\, dP.$$

$\therefore R(\theta, \delta_{\alpha^+}) \leq 2$ if $0 \leq \alpha^+ \leq 1$

Now,

$$\int_{Z>0} Z^2 dP = (2+\eta^2)\ \Phi(\eta/\sqrt{2}) + \sqrt{2}\eta\varphi(\eta/\sqrt{2})$$

and $$\int_{Z>0} Z\, dP = \sqrt{2}\ \varphi(\eta/\sqrt{2}) + \eta\ \Phi(\eta/\sqrt{2}).$$

Writing $\xi = \eta/\sqrt{2}$, we have

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) = 2[(1 - 2\alpha^+(1-\alpha^+)\{(1+\xi^2)\ \Phi(\xi) + \xi\varphi(\xi)\}$$
$$+ 2\xi(1-\alpha^+)\{\varphi(\xi) + \xi\ \Phi(\xi)\}],$$

or

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) = 2[1 + 2(1-\alpha^+)^2\ \xi(\varphi(\xi) + \xi\ \Phi(\xi)) - 2\alpha^+(1-\alpha^+)\ \Phi(\xi)]. \tag{3.3.5}$$

Since $\varphi(\xi) + \xi\ \Phi(\xi) \geq 0$ for all $\underset{\sim}{\xi}$, we have

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) \leq 2 \quad \text{for all } \xi \leq 0 \text{ whenever } 0 \leq \alpha^+ \leq 1.$$

Therefore, $\underline{\delta}_{\alpha^+}$ improves upon $\underline{X}$ whenever $0 \leq \alpha^+ < 1$. From Corollary 3.2.2, we have $\underline{X}$ a minimax estimator and so $\{\underline{\delta}_{\alpha^+}(\underline{X}):$ $0 \leq \alpha^+ \leq 1\}$ is a class of minimax estimators.

Now,

$$2(1-\alpha^+)^2 \; \xi(\varphi(\xi) + \xi \; \Phi(\xi)) - 2\alpha^+(1-\alpha^+) \; \Phi(\xi)$$

$$= 2(1-\alpha^+)[\xi(\varphi(\xi)+\xi \; \Phi(\xi)) - \alpha^+(\Phi(\xi) + \xi(\varphi(\xi) + \xi \; \Phi(\xi)))]$$

$$\geq 0$$

for $\alpha^+ > 1$ as the second term in the square brackets is always positive (a proof of this fact is contained in the proof of the next theorem). Hence $\underline{\delta}_{\alpha^+}$, for $\alpha^+ > 1$, is improved by $\underline{X}$ itself. However, $\underline{\delta}_{\alpha^+}$, for $\alpha^+ < 0$, has risk more than 2 at $\theta_1 = \theta_2$ and less than 2 for θ_2 much larger than θ_1. Therefore they are not comparable to $\underline{X}$.

We also note that for $\alpha^+ > \frac{1}{2}$, $\underline{\delta}_{\alpha^+}$ is inadmissible by Lemma 3.3.1 and so we can restrict attention to $\alpha^+ \leq \frac{1}{2}$.

Next, we use the Brewster-Zidek technique (1974), discussed in Chapter 2, to find a class of estimators admissible in the class of mixed estimators.

<u>Theorem 3.3.3</u>: The estimator $\underline{\delta}_{\alpha^+}$, for $\alpha^+ \leq \frac{1}{2}$, is admissible among the mixed estimators.

<u>Proof</u>: From (3.3.5) we see that the risk function of $\underline{\delta}_{\alpha^+}$ for fixed ξ, is a convex function of α^+. Differentiating $R(\underline{\theta},\underline{\delta}_{\alpha^+})$ with respect to α^+ and equating to zero gives

$$\alpha^+(\xi) = 1 - \frac{1}{2} f^*(\xi), \tag{3.3.6}$$

where $$f^*(\xi) = \frac{\Phi(\xi)}{(1+\xi^2) \; \Phi(\xi) + \xi \; \varphi(\xi)} . \tag{3.3.7}$$

$\alpha^+(\xi)$ minimizes $R(\underline{\theta},\underline{\delta}_{\alpha^+})$ at ξ. Write

$$g(\xi) = (1+\xi^2)\ \Phi(\xi) + \xi\,\varphi(\xi).$$

Clearly $g(0) = \frac{1}{2}$, $\lim\limits_{\xi\to-\infty} g(\xi) = 0$, $\lim\limits_{\xi\to+\infty} g(\xi) = +\infty$.

Also $g'(\xi) = 2(\varphi(\xi) + \xi\ \Phi(\xi)) \geq 0$ for all ξ. Hence $g(\xi)$ is an increasing function of ξ and $g(\xi) \geq 0$ for all ξ.

Since $0 < g(\xi) \leq \Phi(\xi)$ for $\xi \leq 0$, we have $f^*(\xi) \geq 1$ for $\xi \leq 0$ so that $\alpha^+(\xi)$ is bounded above by $\frac{1}{2}$ in $\xi \leq 0$. The upper bound is attained at $\xi = 0$.

Therefore, we conclude by the Brewster-Zidek technique that the estimator $\underline{\delta}_{\alpha^+}$ for $\alpha^+ \in (\frac{1}{2}, 1]$ is improved by $\underline{\delta}_{1/2}$.

By a repeated use of L'Hospital's rule, we get

$$\begin{aligned}\lim_{\xi\to-\infty} f^*(\xi) &= \lim_{\xi\to-\infty}\frac{\Phi(\xi)}{g(\xi)} = \lim_{\xi\to-\infty}\frac{\varphi(\xi)}{g'(\xi)}\\ &= \lim_{\xi\to-\infty}\frac{-\xi\varphi(\xi)}{2\Phi(\xi)}\\ &= \lim_{\xi\to-\infty}\frac{1}{2}(\xi^2-1)\\ &= +\infty\end{aligned}$$

and so $\lim\limits_{\xi\to-\infty} \alpha^+(\xi) = -\infty$. Now, we have $\alpha^+(\xi)$ continuous, $\alpha^+(0) = \frac{1}{2}$ and $\alpha^+(\xi) \leq \frac{1}{2}$ for $\xi \leq 0$, therefore any value a in the interval $(-\infty,\frac{1}{2}]$ is an $\alpha^+(\xi)$ and so minimizes $R(\underline{\theta},\underline{\delta}_{\alpha^+})$ at some ξ. Hence the estimator $\underline{\delta}_{\alpha^+}(\underline{X})$ for $\alpha^+ \leq \frac{1}{2}$ is admissible among the mixed estimators (3.3.3).

<u>Remark 3.3.4</u>: It can be shown easily that $\underline{\delta}_{1/2}(\underline{X})$ is also the MLE of $\underline{\theta}$ for $\theta_1 \leq \theta_2$. Finding the MLE, here, is equivalent to minimizing $(x_1-\theta_1)^2 + (x_2-\theta_2)^2$ with respect to (θ_1,θ_2) subject to

the condition that $\theta_1 \le \theta_2$. When $x_1 \le x_2$, the minimizing choice of (θ_1,θ_2) is, clearly, (x_1,x_2). When $x_1 > x_2$, the minimizing (θ_1,θ_2) is $(\frac{x_1+x_2}{2}, \frac{x_1+x_2}{2})$. To prove this, we show that

$$\tfrac{1}{2}(x_1-x_2)^2 \le (x_1-\theta_1)^2 + (x_2-\theta_2)^2$$

for all $x_1 > x_2$ and $\theta_1 \le \theta_2$ or, equivalently

$$\frac{1}{\sqrt{2}}(x_1-x_2) \le [(x_1-\theta_1)^2 + (x_2-\theta_2)^2]^{1/2} \tag{3.3.8}$$

for all $x_1 > x_2$ and $\theta_1 \le \theta_2$. The right hand side of (3.3.8) is the Euclidean distance between the points (x_1,x_2) and (θ_1,θ_2) and is larger than $\frac{x_1-x_2}{2}$, the length of the perpendicular from (x_1,x_2) on the line $x_1 = x_2$.

This proves our claim.

Remark 3.3.5: Consider the estimators

$$\underline{\delta}_o^*(\underline{X}) = \begin{pmatrix} \mathrm{Min}(X_1,X_2) \\ X_2 \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \text{if } X_1 \le X_2$$
$$= \begin{pmatrix} X_2 \\ X_2 \end{pmatrix}, \quad \text{if } X_1 > X_2 \tag{3.3.9}$$

and
$$\underline{\delta}_1^*(\underline{X}) = \begin{pmatrix} X_1 \\ \mathrm{Max}(X_1,X_2) \end{pmatrix} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \text{if } X_1 \le X_2$$
$$= \begin{pmatrix} X_1 \\ X_1 \end{pmatrix}, \quad \text{if } X_1 > X_2 . \tag{3.3.10}$$

The risk of $\underline{\delta}_o^*$ is

$$R(\underline{\theta},\underline{\delta}_o^*) = \int_{X_1 \le X_2} (X_1-\theta_1)^2 dP + \int_{X_1 \le X_2} (X_2-\theta_2)^2 dP$$

$$+ \int_{X_1 > X_2} (X_2-\theta_1)^2 dP + \int_{X_1 > X_2} (X_2-\theta_2)^2 dP \; . \qquad (3.3.11)$$

The second and fourth integrals on the right hand side of (3.3.11) can be combined to get simply $E(X_2-\theta_2)^2$ which is unity and so

$$R(\underline{\theta},\underline{\delta}_o^*) = 1 + \int_{X_1 \le X_2} (X_1-\theta_1)^2 dP + \int_{X_1 > X_2} (X_2-\theta_1)^2 dP \; . \qquad (3.3.12)$$

Now

$$\int_{X_1 \le X_2} (X_1-\theta_1)^2 dP = \int_{-\infty}^{\infty} \int_{-\infty}^{x_2} (x_1-\theta_1)^2 \varphi(x_1-\theta_1) \varphi(x_2-\theta_2) dx_1 dx_2$$

$$= \int_{-\infty}^{\infty} (\theta_1-x_2) \varphi(x_2-\theta_1) \varphi(x_2-\theta_2) dx_2$$

$$+ \int_{-\infty}^{\infty} \Phi(x_2-\theta_1) \varphi(x_2-\theta_2) dx_2 \qquad (3.3.13)$$

using an integration by parts with respect to x_1.

To simplify the integrals in (3.3.13), we use the following two identities:

$$\frac{1}{b_1} \varphi\left(\frac{x-a_1}{b_1}\right) \frac{1}{b_2} \varphi\left(\frac{x-a_2}{b_2}\right)$$

$$= \frac{1}{\sqrt{(b_1^2+b_2^2)}} \varphi\left(\frac{a_1-a_2}{\sqrt{(b_1^2+b_2^2)}}\right) \frac{\sqrt{(b_1^2+b_2^2)}}{b_1 b_2} \varphi\left(\frac{\sqrt{(b_1^2+b_2^2)}}{b_1 b_2} \left(x - \frac{b_2^2 a_1 + b_1^2 a_2}{b_1^2+b_2^2}\right)\right)$$

$$(3.3.14)$$

and
$$\int_{-\infty}^{\infty} \Phi\left(\frac{x-a_1}{b_1}\right) \frac{1}{b_2} \varphi\left(\frac{x-a_2}{b_2}\right) dx = \Phi\left(\frac{a_2-a_1}{\sqrt{(b_1^2+b_2^2)}}\right) \; . \qquad (3.3.15)$$

The proof of (3.3.14) is straightforward. To verify (3.3.15), consider two independent normal random variables Y_1 and Y_2 with

means a_1, a_2 and variances b_1^2, b_2^2 respectively. Then both the sides of (3.3.15) are equal to $P(Y_1 \leq Y_2)$.

Using the above identities, we get from (3.3.13),

$$\int_{X_1 \leq X_2} (X_1-\theta_1)^2 dP = 1 - \Phi(\xi) + \frac{1}{2}\, \xi\, \varphi(\xi) .$$

In a similar way, we calculate

$$\int_{X_1 > X_2} (X_2-\theta_1)^2 dP = (1+2\xi^2)\ \Phi(\xi) + \frac{3}{2}\, \xi\varphi(\xi),$$

and thus (3.3.12) gives

$$R(\underline{\theta},\underline{\delta}_o^*) = 2[1 + \xi(\varphi(\xi) + \xi\ \Phi(\xi))],$$

which is less than or equal to 2 for $\xi \leq 0$. Therefore, $\underline{\delta}_o^*$ dominates $\underline{X}$ and is a minimax estimator by Corollary 3.2.2.

The risk of $\underline{\delta}_1^*$ is

$$R(\underline{\theta},\underline{\delta}_1^*) = \int_{X_1 \leq X_2} (X_1-\theta_1)^2 dP + \int_{X_1 \leq X_2} (X_2-\theta_2)^2 dP$$

$$+ \int_{X_1 > X_2} (X_1-\theta_1)^2 dP + \int_{X_1 > X_2} (X_1-\theta_2)^2 dP .$$

Combining the first and the third integrals in the above expression

$$R(\underline{\theta},\underline{\delta}_1^*) = 1 + \int_{X_1 \leq X_2} (X_2-\theta_2)^2 dP + \int_{X_1 > X_2} (X_1-\theta_2)^2 dP .$$

We notice that the integrals $\int_{X_1 \leq X_2} (X_2-\theta_2)^2 dP$ and $\int_{X_1 > X_2} (X_1-\theta_2)^2 dP$ can be obtained from $\int_{X_1 \leq X_2} (X_1-\theta_1)^2 dP$ and $\int_{X_1 > X_2} (X_2-\theta_1)^2 dP$ of (3.3.12) by an interchange of θ_1 and θ_2 and some adjustment. This gives

$$R(\underline{\theta},\underline{\delta}_1^*) = 2[1 + \xi(\varphi(\xi) + \xi\ \Phi(\xi))] = R(\underline{\theta},\underline{\delta}_o^*) \quad \text{for all } \underline{\theta} \in \Omega_o$$

and so the estimators $\underline{\delta}_1^*$ and $\underline{\delta}_o^*$ are equivalent. Since the loss function is convex, any convex combination

$$\underline{\delta}^*_\beta = \beta\, \underline{\delta}^*_1 + (1-\beta)\, \underline{\delta}^*_o , \qquad 0 < \beta < 1 \tag{3.3.16}$$

of $\underline{\delta}^*_1$ and $\underline{\delta}^*_o$ is an improvement upon $\underline{\delta}^*_o$ and $\underline{\delta}^*_1$. Consequently, $\{\underline{\delta}^*_\beta : 0 \leq \beta \leq 1\}$ is a class of minimax estimators. The risk of $\underline{\delta}^*_\beta$ is

$$R(\underline{\theta}, \underline{\delta}^*_\beta) = \int_{X_1 \leq X_2} (X_1-\theta_1)^2 dP + \int_{X_1 \leq X_2} (X_2-\theta_2)^2 dP + \int_{X_1 > X_2} (\beta X_1 + (1-\beta) X_2 - \theta_1)^2 dP$$

$$+ \int_{X_1 > X_2} (\beta X_1 + (1-\beta) X_2 - \theta_2)^2 \, dP .$$

After some simplification, we get

$$R(\underline{\theta}, \underline{\delta}^*_\beta) = 2 + ((1-\beta)^2 + \beta^2) \int_{Z>0} Z^2 dP + 2(1-\beta) \int_{X_1 > X_2} (X_2-X_1)(X_1-\theta_1) dP$$

$$+ 2\beta \int_{X_1 > X_2} (X_1-X_2)(X_2-\theta_2) dP , \tag{3.3.17}$$

where Z is the same as in (3.3.4). The integral $\int_{Z>0} Z^2 dP$ was calculated in the derivation of $R(\underline{\theta}, \underline{\delta}_{\alpha^+})$ and the last two integrals are calculated using the identities (3.3.14) and (3.3.15) and are seen to have the identical value $-\Phi(\xi)$. This gives

$$R(\underline{\theta}, \underline{\delta}^*_\beta) = 2\left[1 + (\beta^2 + (1-\beta)^2)((1+\xi^2)\ \Phi(\xi) + \xi\, \varphi(\xi)) - \Phi(\xi)\right] . \tag{3.3.18}$$

Now, since $(1+\xi^2)\ \Phi(\xi) + \xi\, \varphi(\xi) > 0$ (see proof of Theorem 3.3.3), $\beta = \frac{1}{2}$ minimizes $R(\underline{\theta}, \underline{\delta}^*_\beta)$ and so $\underline{\delta}^*_{1/2}$ is the best choice in the class $\{\underline{\delta}^*_\beta : 0 \leq \beta \leq 1\}$. Incidentally, $\underline{\delta}^*_{1/2}$ is the same as the mixed estimator $\underline{\delta}_{1/2}$. We also observe that $\underline{\delta}_\beta$ is better than $\underline{\delta}^*_\beta$ and $\underline{\delta}^*_{1-\beta}$ for $0 \leq \beta \leq \frac{1}{2}$ and is worse for $\frac{1}{2} < \beta \leq 1$.

<u>Remark 3.3.6</u>: The problem is invariant under the transformations:

$$\begin{aligned} X_i &\to X_i + a, \\ \theta_i &\to \theta_i + a, \\ \hat{\theta}_i &\to \hat{\theta}_i + a, \quad i = 1, 2. \end{aligned} \tag{3.3.19}$$

The form of an equivalent estimator is

$$\underline{\delta}(\underline{X}) = \begin{bmatrix} \delta_1(\underline{X}) \\ \delta_2(\underline{X}) \end{bmatrix} = \begin{bmatrix} X_1 + \Psi_1(X_1-X_2) \\ X_2 + \Psi_2(X_1-X_2) \end{bmatrix}. \qquad (3.3.20)$$

Clearly, the risk of an equivariant estimator is a function of $\theta_1-\theta_2$ and so there is no best equivariant estimator.

A subclass of equivariant estimators (3.3.20) is

$$\underline{\delta}(\underline{X}) = \begin{bmatrix} X_1 - \Psi(X_1-X_2) \\ X_2 + \Psi(X_1-X_2) \end{bmatrix}. \qquad (3.3.21)$$

Notice that the mixed estimators are equivariant and they are of the form (3.3.21).

An application of Lemma 3.3.1 proves the following result.

Theorem 3.3.4: Let $\underline{\delta}(\underline{X})$ be as in (3.3.21) and $P_{\underline{\theta}'}(\Psi(X_1-X_2) < \frac{X_1-X_2}{2}) > 0$ for some $\underline{\theta}' \in \Omega_o$. Then $\underline{\delta}(\underline{X})$ is inadmissible.

It is interesting to note that Theorem 3.3.4 can also be proved using the Brewster-Zidek technique by improving the estimator $\underline{\delta}(\underline{X})$ on the orbits $\{(X_1,X_2): X_1-X_2 = c\}$, $c \in R^1$. The proof is sketched below:

The risk of the estimator $\underline{\delta}(\underline{X})$ of (3.3.21) is

$$R(\underline{\theta},\underline{\delta}) = E^Z R(\underline{\theta}, \underline{\delta}, Z),$$

where $R(\underline{\theta},\underline{\delta},z) = E\left[\{(X_1-\Psi(Z)-\theta_1)^2 + (X_2+\Psi(Z)-\theta_2)^2\} \mid Z = z\right]$

with Z the same as in (3.3.4).

Now, the conditional distributions of X_1 given $Z = z$ and X_2 given $Z = z$ are $N(\frac{\theta_1+\theta_2+z}{2}, \frac{1}{2})$ and $N(\frac{\theta_1+\theta_2-z}{2}, \frac{1}{2})$ respectively and so for fixed z and $\underline{\theta}$, the value of Ψ minimizing $R(\underline{\theta},\underline{\delta},z)$ is

$$\Psi^*_{\underline{\theta}}(z) = \frac{1}{2}(z + \theta_2 - \theta_1),$$

which has an infimum z/2 over the region $\theta_2 \geq \theta_1$. Since $R(\underline{\theta}, \underline{\delta}, z)$ is, for fixed z and $\underline{\theta}$, convex as a function of Ψ with a minimum at $\Psi^*_{\underline{\theta}}(z) \geq z/2$, it follows that $\Psi(z) \leq z/2$ implies

$$E[\{(X_1 - \Psi(z) - \theta_1)^2 + (X_2 + \Psi(z) - \theta_2)^2\} \mid Z = z]$$

$$\geq E[\{(X_1 - \frac{z}{2} - \theta_1)^2 + (X_2 + \frac{z}{2} - \theta_2)^2\} \mid Z = z]$$

for all $\theta_1 \leq \theta_2$. Thus the estimator $\underline{\delta}^*(\underline{X})$, obtained by replacing $\Psi(Z)$ by $\max(\Psi(Z), \frac{Z}{2})$ in $\underline{\delta}(\underline{X})$, dominates $\underline{\delta}(\underline{X})$, provided $P_{\underline{\theta}'}(\Psi(Z) < Z/2)$ is positive for some $\underline{\theta}' \in \Omega_o$. This completes the proof of Theorem 3.3.4.

Inadmissibility of Mixed Estimators

We prove below the inadmissibility of the mixed estimators.

Theorem 3.3.5: Mixed estimator $\underline{\delta}_{\alpha^+}$, for any real α^+, is inadmissible.

Proof: In view of Theorem 3.3.3, it suffices to consider $\underline{\delta}_{\alpha^+}$ for $\alpha^+ \leq \frac{1}{2}$. Since the class of generalized Bayes estimators is complete (see p. 43, Chapter 2 for the results of Caridi (1983), Berger and Srinivasan (1978) and Brown (1971)), it is enough to show that $\underline{\delta}_{\alpha^+}$ is not generalized Bayes.

Now,

$$\underline{\delta}_{\alpha^+} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, \quad \text{if } X_1 \leq X_2$$

$$= \begin{pmatrix} \alpha X_1 + (1-\alpha)X_2 \\ (1-\alpha)X_1 + \alpha X_2 \end{pmatrix}, \quad \text{if } X_1 > X_2 .$$

Suppose $\underline{\delta}_{\alpha^+}$ is generalized Bayes for the prior $f(\underline{\theta})$ with respect to Lebesgue measure. Then

$$\frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \underline{\theta}\ \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})d\theta_1 d\theta_2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})\ d\theta_1 d\theta_2} = \underline{x} \quad \text{for all } x_1 \leq x_2.$$

This implies that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\underline{\theta}-\underline{x})\ \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})\ d\theta_1 d\theta_2 = \underline{0} \quad \text{for all } x_1 \leq x_2,$$

or equivalently

$$\frac{d}{dx_1}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})d\theta_1 d\theta_2 = 0 \quad \text{for all } x_1 \leq x_2,$$

and

$$\frac{d}{dx_2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})d\theta_1 d\theta_2 = 0 \quad \text{for all } x_1 \leq x_2.$$

Therefore, we must have

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ f(\underline{\theta})d\theta_1 d\theta_2 = K \text{ (a constant)} \quad \text{for all } x_1 \leq x_2.$$

Now, the family of bivariate normal distributions $\{N_2(\underline{x}, I): x_1 \leq x_2\}$ is complete (see, for example, Lehmann, 1959, p. 132). Hence, we must have $f(\underline{\theta}) = K$ a.e. Lebesgue and so $\underline{\delta}_{\alpha^+}(\underline{x}) = \underline{x}$ for all $\underline{x}$, which is a contradiction.

This completes the proof of the theorem.

<u>Remark 3.3.7</u>: The above method of proving inadmissibility does not yield a better estimator. We do not know what estimator is better than $\underline{\delta}_{\alpha^+}$ for $\alpha^+ \leq \frac{1}{2}$.

Remark 3.3.8: Katz (1963) considered the Pitman estimator $\underline{\delta}_p$ of (3.2.8), which in this case is

$$\underline{\delta}_p(\underline{X}) = \begin{bmatrix} X_1 - \frac{1}{\sqrt{2}}\,\nu\left(\frac{X_2-X_1}{\sqrt{2}}\right) \\ X_2 + \frac{1}{\sqrt{2}}\,\nu\left(\frac{X_2-X_1}{\sqrt{2}}\right) \end{bmatrix},$$

where $\nu(t) = \varphi(t)/\Phi(t)$. He showed that $\underline{\delta}_p(\underline{X})$ is admissible and minimax, though his proofs are inadequate. Blumenthal and Cohen (1968b) show in general for distributions with location parameter densities that $\underline{\delta}_p$ of (3.2.8) is both admissible and minimax under certain conditions. It is tempting to compare $\underline{\delta}_p$ with the mixed estimator. The risk function of $\underline{\delta}_p$, as given by Katz, is

$$R(\underline{\theta},\underline{\delta}_p) = 2 + \xi\, E\, \nu\left(\frac{X_2-X_1}{\sqrt{2}}\right).$$

Since ν is a nonnegative function, $R(\underline{\theta},\underline{\delta}_p) \leq 2$ for all $\xi \leq 0$. When $\xi = 0$, that is, $\theta_1 = \theta_2$, $R(\underline{\theta},\underline{\delta}_p) = 2$ whereas $R(\underline{\theta},\underline{\delta}_{\alpha^+}) = 2[1-\alpha^+(1-\alpha^+)]$ which is always less than 2 whenever $0 < \alpha^+ < 1$. The improvement is maximum when $\alpha^+ = \frac{1}{2}$. Therefore, $\underline{\delta}_{\alpha^+}$ and $\underline{\delta}_p$ are not comparable for $0 < \alpha^+ < 1$. $\underline{\delta}_{1/2}$ will be preferred when θ_1 and θ_2 are known to be quite close. When the difference is large, $\underline{\delta}_p$ will be preferred.

3.4 Two Normal Populations with Unequal Variances

Katz (1963) has considered the normal problem for equal and known variances, whereas Blumenthal and Cohen (1968b) take up the estimation of location parameters when the scale parameters are equal and known. In this section, we study two normal populations with unequal known variances and show that the MLE as well as the mixed estimators remain minimax. Unlike the equal variances case,

the MLE is not a mixed estimator and it dominates some of the mixed estimators including $\underline{\delta}_{1/2}(\underline{X})$. The Pitman estimator $\underline{\delta}_p$ is admissible (see Cohen and Sackrowitz (1970)). We prove its minimaxity and compare it with the MLE and the mixed estimators.

Let X_1 and X_2 be independent normal random variables with means θ_1 and θ_2 and variances τ and 1 respectively. We assume a priori that $\theta_1 \leq \theta_2$. We want to estimate $\underline{\theta} = (\theta_1, \theta_2)$ with squared error loss.

The Mixed Estimators

We have seen in Lemma 3.3.1 that the mixed estimators (3.3.3) improve upon the usual estimator $\underline{X}$. The exact risk expression is:

$$R(\underline{\theta}, \underline{\delta}_{\alpha^+}) = (\tau+1)\left[1 + 2(1-\alpha^+)^2 \lambda(\varphi(\lambda) + \lambda\ \Phi(\lambda)) - 2\alpha^+(1-\alpha^+)\ \Phi(\lambda)\right], \tag{3.4.1}$$

where $\lambda = (\theta_1 - \theta_2)/\sqrt{(\tau+1)}$.

Clearly, $R(\underline{\theta}, \underline{\delta}_{\alpha^+}) \leq \tau + 1$ for $\lambda \leq 0$, $0 \leq \alpha^+ < 1$. Thus, $\underline{\delta}_{\alpha^+}$, for $0 \leq \alpha^+ < 1$, improves upon $\underline{X}$ and so is minimax by Corollary 3.2.2. The estimator $\underline{\delta}_{\alpha^+}$ for $\alpha^+ > 1$ is improved by $\underline{X}$ itself. Also the estimator $\underline{\delta}_{\alpha^+}$, $\alpha^+ < 0$ is not comparable to $\underline{X}$.

Using the Brewster-Zidek technique, as in the proof of Theorem 3.3.3, we have the following result.

<u>Theorem 3.4.1</u>: The estimator $\underline{\delta}_{\alpha^+}$ for $\alpha^+ \leq \frac{1}{2}$ is admissible among the mixed estimators.

The Maximum Likelihood Estimator

For this problem, finding the MLE is equivalent to minimizing $\frac{1}{\tau}(x_1-\theta_1)^2 + (x_2-\theta_2)^2$ with respect to (θ_1, θ_2) subject to $\theta_1 \leq \theta_2$.

This gives $\underline{\delta}^*_{1/(\tau+1)}(\underline{X})$ defined by (3.3.16) as the MLE. The risk function of $\underline{\delta}^*_\beta$ of (3.3.16) is

$$R(\underline{\theta},\underline{\delta}^*_\beta) = \int_{X_1\le X_2}(X_1-\theta_1)^2 dP + \int_{X_1\le X_2}(X_2-\theta_2)^2 dP$$

$$+ \int_{X_1>X_2}(\beta X_1+(1-\beta)X_2-\theta_1)^2 dP + \int_{X_1>X_2}(\beta X_1+(1-\beta)X_2-\theta_2)^2 dP,$$

which after some simplification becomes

$$R(\underline{\theta},\underline{\delta}^*_\beta) = (1+\tau) + (\beta^2+(1-\beta)^2)\int_{X_1>X_2}(X_1-X_2)^2 dP$$

$$+ 2\beta\int_{X_1>X_2}(X_1-X_2)(X_2-\theta_2)dP + 2(1-\beta)\int_{X_1>X_2}(X_2-X_1)(X_1-\theta_1)dP.$$

The integrals in the above expression are calculated to be

$$\int_{X_1>X_2}(X_1-X_2)^2 dP = (1+\tau)\ \{(1+\lambda^2)\ \Phi(\lambda) + \lambda\ \varphi(\lambda)\},$$

$$\int_{X_1>X_2}(X_1-X_2)(X_2-\theta_2)dP = -\ \Phi(\lambda), \quad \text{and}$$

$$\int_{X_1>X_2}(X_2-X_1)(X_1-\theta_1)dP = -\ \tau\ \Phi(\lambda),$$

where λ is the same as in (3.4.1). Thus, we have

$$R(\underline{\theta},\underline{\delta}^*_\beta) = (1+\tau)\left[1 + (\beta^2+(1-\beta)^2)\ \{(1+\lambda^2)\ \Phi(\lambda) + \lambda\ \varphi(\lambda)\}\right]$$

$$- 2((1-\beta)\tau + \beta)\ \Phi(\lambda). \tag{3.4.2}$$

Now we prove the following theorem.

<u>Theorem 3.4.2</u>: (i) The estimator $\underline{\delta}_{\alpha^+}$ for $\alpha^+ > 1 - \frac{1}{(\tau+1)}\ \frac{\sqrt{(\tau^2+1)}}{\sqrt{2}}$, is dominated by the MLE. (ii) The MLE is admissible among the mixed estimators and (iii) The MLE is also admissible in the class $\{\underline{\delta}^*_\beta:\ 0 \le \beta \le 1\}$.

<u>Proof</u>: From (3.4.2), the risk of $\underline{\delta}^*_{1/(\tau+1)}$ is

$$R(\underline{\theta},\underline{\delta}^*_{1/(\tau+1)}) = (\tau+1)\left[1+\frac{(\tau^2+1)}{(\tau+1)^2}\{\lambda(\varphi(\lambda)+\lambda\,\Phi(\lambda))-\Phi(\lambda)\}\right]. \tag{3.4.3}$$

We notice that $\lambda(\varphi(\lambda)+\lambda\,\Phi(\lambda)) \le 0$ for $\lambda \le 0$ and $2\alpha^+(1-\alpha^+) \le \frac{1}{2} < \frac{(\tau^2+1)}{(\tau+1)^2}$ for α^+ and $\tau \ne 1$. Now (3.4.1) and the fact that $2(1-\alpha^+)^2 \le \frac{(\tau^2+1)}{(\tau+1)^2}$ prove (i). The assertion (ii) follows from the fact that at $\theta_1 = \theta_2$ the risk of $\underline{\delta}_{\alpha^+}$ is always larger than the risk of the MLE. To prove (iii), consider the risk function of $\underline{\delta}^*_\beta$ and find that the β minimizing the $R(\underline{\theta},\underline{\delta}^*_\beta)$ for fixed $\underline{\theta}$ is

$$\beta(\lambda) = \frac{1}{2}\left[1+\frac{1-\tau}{1+\tau}f^*(\lambda)\right],$$

where f^* is defined by (3.3.7). Clearly $\beta = \frac{1}{1+\tau}$ when $\lambda = 0$ and so $\underline{\delta}^*_{1/(\tau+1)}$ minimizes $R(\underline{\theta},\underline{\delta}^*_\beta)$ at $\theta_1 = \theta_2$. Convexity of the risk function in β ensures that it is the unique minimizer. Hence $\underline{\delta}^*_{1/(\tau+1)}$ is admissible among the estimators $\underline{\delta}^*_\beta$, $0 \le \beta \le 1$.

<u>Remark 3.4.1</u>: It was shown in the proof of Theorem 3.3.3 that $\inf_{\lambda\le 0} f^*(\lambda) = 1$ and $\sup_{\lambda\le 0} f^*(\lambda) = +\infty$. Consequently, $\inf_{\lambda\le 0}\beta(\lambda) = \frac{1}{1+\tau}$, $\sup_{\lambda\le 0}\beta(\lambda) = +\infty$, if $\tau < 1$ and $\inf_{\lambda\le 0}\beta(\lambda) = -\infty$, $\sup_{\lambda\le 0}\beta(\lambda) = \frac{1}{1+\tau}$, if $\tau > 1$. The Brewster-Zidek technique (1974) then proves that

(i) if $\tau < 1$, $\underline{\delta}^*_\beta$ for $\beta \ge \frac{1}{\tau+1}$ is admissible among $\underline{\delta}^*_\beta$'s, and

(ii) if $\tau > 1$, $\underline{\delta}^*_\beta$ for $\beta \le \frac{1}{\tau+1}$ is admissible among $\underline{\delta}^*_\beta$'s.

<u>Remark 3.4.2</u>: Inadmissibility of the mixed estimators can be proved following the same lines as in the proof of Theorem 3.3.5.

Remark 3.4.3: The MLE is also inadmissible, since as in the proof of Theorem 3.3.5, it can be shown to be not generalized Bayes.

Invariance

We can introduce invariance in this problem the same way as in Section 3.3. The form of an equivariant estimator is the same as in (3.3.20). By an application of Lemma 3.3.1 or the orbit by orbit improvement technique of Brewster and Zidek (1974) we see that Theorem 3.3.4 holds also in the set up of unequal variances.

Next, we consider the Pitman estimator $\underline{\delta}_p$ of (3.2.8) which in this case is given by

$$\underline{\delta}_p(\underline{X}) = \begin{bmatrix} \delta_{p1}(\underline{X}) \\ \delta_{p2}(\underline{X}) \end{bmatrix}, \tag{3.4.4}$$

where

$$\delta_{p1}(x) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_2} \theta_1 \frac{1}{\sqrt{\tau}}\varphi(\frac{x_1-\theta_1}{\sqrt{\tau}})\ \varphi(x_2-\theta_2)d\theta_1 d\theta_2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_2} \frac{1}{\sqrt{\tau}}\varphi(\frac{x_1-\theta_1}{\sqrt{\tau}})\ \varphi(x_2-\theta_2)\ d\theta_1 d\theta_2}$$

$$= x_1 - \frac{\tau}{\sqrt{1+\tau}}\ \nu(\frac{x_2-x_1}{\sqrt{1+\tau}}) \tag{3.4.5}$$

and

$$\delta_{p2}(\underline{x}) = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_2} \theta_2 \frac{1}{\sqrt{\tau}}\varphi(\frac{x_1-\theta_1}{\sqrt{\tau}})\ \varphi(x_2-\theta_2)d\theta_1 d\theta_2}{\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_2} \frac{1}{\sqrt{\tau}}\varphi(\frac{x_1-\theta_1}{\sqrt{\tau}})\ \varphi(x_2-\theta_2)d\theta_1 d\theta_2}$$

$$= x_2 + \frac{1}{\sqrt{1+\tau}}\ \nu(\frac{x_2-x_1}{\sqrt{1+\tau}}), \tag{3.4.6}$$

ν being the same as in Remark 3.3.8.

We show in the following theorem that $\underline{\delta}_p$ is minimax.

Theorem 3.4.3: The estimator $\underline{\delta}_p$ of $\underline{\theta}$, $\theta_1 \le \theta_2$, is minimax when the loss function is squared error.

Proof: The risk of $\underline{\delta}_p$ is

$$R(\underline{\theta},\underline{\delta}_p) = E\,|\,\underline{\delta}_p-\underline{\theta}\,|^2$$

$$= E\left[X_1 - \frac{\tau}{\sqrt{1+\tau}}\,\nu\left(\frac{X_2-X_1}{\sqrt{1+\tau}}\right) - \theta_1\right]^2 + E\left[X_2 + \frac{1}{\sqrt{1+\tau}}\,\nu\left(\frac{X_2-X_1}{\sqrt{1+\tau}}\right) - \theta_2\right]^2$$

$$= 1 + \tau + \frac{\tau^2+1}{\tau+1}\,E\,\nu^2(W) - \frac{2\tau}{\sqrt{1+\tau}}\,E(X_1-\theta_1)\,\nu(W)$$

$$+ \frac{2}{\sqrt{1+\tau}}\,E(X_2-\theta_2)\,\nu(W), \tag{3.4.7}$$

where $W = \dfrac{X_2-X_1}{\sqrt{1+\tau}} \sim N(\zeta,1)$ with $\zeta = \dfrac{\theta_2-\theta_1}{\sqrt{1+\tau}}$ nonnegative. Now,

$$E\,\nu^2(W) = \int_{-\infty}^{\infty} \nu^2(w)\,\varphi(w-\zeta)\,dw$$

$$= \int_{-\infty}^{\infty} \varphi(w)\,\frac{\varphi(w)\,\varphi(w-\zeta)}{\Phi^2(w)}\,dw$$

$$= \Phi(w)\,\frac{\varphi(w)\,\varphi(w-\zeta)}{\Phi^2(w)}\Bigg|_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \Phi(w)\left[\frac{1}{\Phi^2(w)}\{-\,w\,\varphi(w)\,\varphi(w-\zeta)\right.$$

$$\left. -\,(w-\zeta)\,\varphi(w)\,\varphi(w-\zeta)\} - \frac{2}{\Phi^3(w)}\,\varphi^2(w)\,\varphi(w-\zeta)\right]dw$$

$$= EW\,\nu(W) + E(W-\zeta)\,\nu(W) + 2E\,\nu^2(W)$$

and so

$$E\,\nu^2(W) = E(\zeta-2W)\,\nu(W). \tag{3.4.8}$$

Substituting for $E\,\nu^2(W)$ from (3.4.8) into (3.4.7) we get

$$R(\underline{\theta},\underline{\delta}_p) = (1+\tau) + \frac{\tau^2+1}{\tau+1} E(\zeta-2W)\ \nu(W) - \frac{2\tau}{\sqrt{1+\tau}} (X_1-\theta_1)\ \nu(W)$$

$$+ \frac{2}{\sqrt{1+\tau}} E(X_2-\theta_2)\ \nu(W)$$

$$= (1+\tau) + \frac{1}{(1+\tau)^{3/2}} E[(\tau^2-2\tau-1)\theta_2 + (\tau^2+2\tau-1)\theta_1$$

$$+ 2(1-\tau)(\tau X_2+X_1)]\ \nu(W) .$$

Since τX_2+X_1 and W are independently distributed, we get

$$R(\underline{\theta},\underline{\delta}_p) = (1+\tau) + \frac{\tau^2+1}{(\tau+1)^{3/2}} (\theta_1-\theta_2)\ E\ \nu(W),$$

which is less than or equal to $(1+\tau)$ for all $\theta_1 \leq \theta_2$. By Corollary 3.2.2, we get the result.

Remark 3.4.4: At $\theta_1 = \theta_2$ the risk of $\underline{\delta}_{\alpha^+}$, $0 < \alpha^+ < 1$, is smaller than the risk of $\underline{\delta}_p$. Since $\underline{\delta}_p$ is admissible, the two estimators are not comparable. Also the risk of the MLE is smaller than that of $\underline{\delta}_p$ at $\theta_1 = \theta_2$. Therefore, both the MLE and the mixed estimator $\underline{\delta}_{\alpha^+}$, $0 < \alpha^+ < 1$, though inadmissible cannot be improved by $\underline{\delta}_p$. We do not know yet any estimator dominating the MLE.

Remark 3.4.5: Observe that the sample means are sufficient and so the case of samples of different sizes from the two populations with equal or unequal variances is covered by the results of this section.

3.5 Two Populations with Location Parameter Densities Not Necessarily of the Same Form

Blumenthal and Cohen (1968b) considered estimation of parameters θ_1 and θ_2 of two populations having densities $f(x-\theta_1)$ and $f(x-\theta_2)$ respectively and obtained sufficient conditions for the

minimaxity of the Pitman estimator $\underline{\delta}_p$. In Section 3.4, we proved that the Pitman estimator is minimax when the two populations are normal but have unequal variances. A more general set up would have been of densities $\frac{1}{\sigma_1} f(\frac{x-\theta_1}{\sigma_1})$ and $\frac{1}{\sigma_2} f(\frac{x-\theta_2}{\sigma_2})$, $\sigma_1 \neq \sigma_2$. In this section, we take up densities $f_1(x-\theta_1)$ and $f_2(x-\theta_2)$ with the set up of Section 3.2 and develop sufficient conditions for the minimaxity of the Pitman estimator.

In view of Corollary 3.2.2, to prove the minimaxity of the Pitman estimator $\underline{\delta}_p$, it suffices to show that $R(\underline{\theta},\underline{\delta}_p) \leq R$ for all $\theta_1 \leq \theta_2$. In Theorems 3.5.1 and 3.5.5 we give sufficient conditions under which this holds.

Theorem 3.5.1: If the densities $p_i(x,\underline{y}_i)$, $i = 1,2$, defined in (3.2.4), have finite variance and if

$$\int_{-\infty}^{\infty} u\, p_1(u-v,\underline{y}_1)\, p_2(u+v,\underline{y}_2)\,du = 0 \quad \text{for all } v,\ \underline{y}_1 \text{ and } \underline{y}_2, \tag{3.5.1}$$

then $R(\underline{\theta},\underline{\delta}_p) \leq R$ for all $\theta_1 \leq \theta_2$.

Proof: The proof of the theorem is similar to that of Theorem 3.2 of Blumenthal and Cohen (1968b). Some changes are necessary as we have two different densities $f_1(x-\theta_1)$ and $f_2(x-\theta_2)$.

The risk of the Pitman estimator $\underline{\delta}_p = (\delta_{p1},\delta_{p2})$, by (3.2.7), can be written as

$$R(\underline{\theta},\underline{\delta}_p) = \int_{\mathcal{Y}} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} [(\delta_{p1}-\theta_1)^2 + (\delta_{p2}-\theta_2)^2]\, p_1(x_1-\theta_1,\underline{y}_1)\, p_2(x_2-\theta_2,\underline{y}_2)\,dx_1 dx_2\ \nu(d\underline{y}).$$

If we use variables $Z_1 = \frac{X_1+X_2}{2}$ and $Z_2 = \frac{X_2-X_1}{2}$ in place of X_1, X_2 and define

$$\mu_1 = \frac{\theta_1+\theta_2}{2}, \quad \mu_2 = \frac{\theta_2-\theta_1}{2}, \tag{3.5.2}$$

$$\gamma_1 = \frac{\delta_{p1}+\delta_{p2}}{2}, \quad \gamma_2 = \frac{\delta_{p2}-\delta_{p1}}{2},$$

the risk becomes

$$R(\underline{\theta},\underline{\delta}_p) = 4\left[R_1(\mu_1,\mu_2,\gamma_1) + R_2(\mu_1,\mu_2,\gamma_2)\right], \tag{3.5.3}$$

where

$$R_1(\mu_1,\mu_2,\gamma_1) = \int_{\mathcal{Y}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\gamma_1-\mu_1)^2\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)\,dz_1 dz_2\ \nu(d\underline{y}) \tag{3.5.4}$$

and

$$R_2(\mu_1,\mu_2,\gamma_2) = \int_{\mathcal{Y}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\gamma_2-\mu_2)^2\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)\,dz_1 dz_2\ \nu(d\underline{y}). \tag{3.5.5}$$

Using (3.5.2) in the expression (3.2.8) for δ_{p1} and δ_{p2}, we get

$$\gamma_1 = \frac{\int_0^{\infty}\int_{-\infty}^{\infty}\mu_1\, p_1((Z_1-\mu_1)-(Z_2-\mu_2),\underline{Y}_1)\ p_2((Z_1-\mu_1)+(Z_2-\mu_2),\underline{Y}_2)\,d\mu_1 d\mu_2}{\int_0^{\infty}\int_{-\infty}^{\infty}p_1((Z_1-\mu_1)-(Z_2-\mu_2),\underline{Y}_1)\ p_2((Z_1-\mu_1)+(Z_2-\mu_2),\underline{Y}_2)\,d\mu_1 d\mu_2} \tag{3.5.6}$$

and

$$\gamma_2 = \frac{\int_0^{\infty}\int_{-\infty}^{\infty}\mu_2\, p_1((Z_1-\mu_1)-(Z_2-\mu_2),\underline{Y}_1)\ p_2((Z_1-\mu_1)+(Z_2-\mu_2),\underline{Y}_2)\,d\mu_1 d\mu_2}{\int_0^{\infty}\int_{-\infty}^{\infty}p_1((Z_1-\mu_1)-(Z_2-\mu_2),\underline{Y}_1)\ p_2((Z_1-\mu_1)+(Z_2-\mu_2),\underline{Y}_2)\,d\mu_1 d\mu_2} \tag{3.5.7}$$

provided the integral in the denominator is not zero.

Making another change of variables $u = Z_1-\mu_1$ and $v = Z_2-\mu_2$ in (3.5.6) and (3.5.7) we get

$$\gamma_1 = Z_1 - \frac{\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} u\, p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv}{\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv} \tag{3.5.8}$$

and

$$\gamma_2 = Z_2 - \frac{\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} v\, p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv}{\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv} \tag{3.5.9}$$

provided $\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv > 0.$

Writing

$$\int_{-\infty}^{\infty} u\, p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du = g(v,\underline{Y}),$$

we have

$$\gamma_1 = Z_1 - \frac{\int_{-\infty}^{Z_2} \int_{-\infty}^{\infty} u\, p_1(u-v,\underline{Y}_1)\, p_2(u+v,\underline{Y}_2)\, du\, dv}{\int_{-\infty}^{Z_2} g(v,\underline{Y})\, dv} \tag{3.5.10}$$

and

$$\gamma_2 = Z_2 - \frac{\int_{-\infty}^{Z_2} v\, g(v,\underline{Y})\, dv}{\int_{-\infty}^{Z_2} g(v,\underline{Y})\, dv} \tag{3.5.11}$$

provided $\int_{-\infty}^{Z_2} g(v,\underline{Y})\, dv > 0.$

We notice that γ_2 depends only on $(Z_2,\underline{Y})$ and not on Z_1. Consequently,

$$R_2(\mu_1,\mu_2,\gamma_2) = \int_{\mathcal{Y}} \int_{-\infty}^{\infty} (\gamma_2-\mu_2)^2\, g(z_2-\mu_2,\underline{y})\, dz_2\, \nu(d\underline{y}).$$

Since the density of Z_2 is $2g(z_2-\mu_2,\underline{y})$, $\mu_2 \geq 0$, the estimator γ_2 is simply the generalized Bayes estimator of μ_2 with respect to the uniform prior on the space $\mu_2 \geq 0$ and so

$$R_2(\mu_1,\mu_2,\gamma_2) = \frac{1}{2} R(\mu_2,\gamma_2). \tag{3.5.12}$$

The result of Farrell (1964, p. 980) proves that

$$R(\mu_2,\gamma_2) \leq R(\mu_2,Z_2) = \frac{R}{4} \quad \text{for all } \mu_2 \geq 0 \tag{3.5.13}$$

and $$\lim_{\mu_2\to\infty} R(\mu_2,\gamma_2) = \frac{R}{4}\,. \tag{3.5.14}$$

If the condition (3.5.1) is satisfied we have $\gamma_1 = Z_1$ and $R_1(\mu_1,\mu_2,\gamma_1) = \frac{R}{8}$. This along with (3.5.13) and (3.5.3) proves that

$$R(\underline{\theta},\underline{\delta}_p) \leq R \quad \text{for all } \mu_1 \in R^1 \text{ and } \mu_2 \geq 0.$$

This completes the proof of the theorem.

Some more notation is needed for deriving another sufficient condition for the minimaxity of the Pitman estimator.

Let

$$P_i(z_2,\underline{y}_i) = \int_{-\infty}^{z_2} p_i(x,\underline{y}_i)dx\,, \quad i = 1,2$$

$$G(z_2,\underline{y}) = \int_{-\infty}^{z_2} g(v,\underline{y})dv\,,$$

$$E(z_2,\underline{y}) = \frac{1}{4}\int_{-\infty}^{\infty} x\, p_2(x,\underline{y}_2)\, P_1(x-2z_2,\underline{y}_1)dx$$

$$= -\frac{1}{4}\int_{-\infty}^{\infty} x\, p_2(x,\underline{y}_2)\, [1-P_1(x-2z_2,\underline{y}_1)]\, dx\,,$$

$$F(z_2,\underline{y}) = \frac{1}{4}\int_{-\infty}^{\infty} x\, p_1(x,\underline{y}_1)\, P_2(x+2z_2,\underline{y}_2)dx,$$

$$\hat{E}(z_2,\underline{y}) = E(z_2,\underline{y})/G(z_2,\underline{y}), \tag{3.5.15}$$

$$\hat{F}(z_2,\underline{y}) = F(z_2,\underline{y})/G(z_2,\underline{y}),$$

$$K(z_2,\underline{y}) = -\,(\hat{F}(z_2,\underline{y}) + \hat{E}(z_2,\underline{y})),$$

$$H(z_2,\underline{y}) = \hat{F}(z_2,\underline{y}) - \hat{E}(z_2,\underline{y}),$$

and $$T(z_2,\underline{y}) = \int_{-\infty}^{\infty} x\, p_1(x-z_2,\underline{y}_1)\, p_2(x+z_2,\underline{y}_2)dx$$

We see that

$$\int_{-\infty}^{\infty}\int_{-\infty}^{z_2} u\, p_1(u-v,\underline{y}_1)\, p_2(u+v,\underline{y}_2)dv\, du$$

$$= \frac{1}{4}\int_{-\infty}^{\infty}\int_{-\infty}^{s+2z_2} (s+t)\, p_1(s,\underline{y}_1)\, p_2(t,\underline{y}_2)dt\, ds$$

$$= \frac{1}{4}\Big[\int_{-\infty}^{\infty}\int_{-\infty}^{s+2z_2} s\, p_1(s,\underline{y}_1)\, p_2(t,\underline{y}_2)dt\, ds$$

$$+ \int_{-\infty}^{\infty}\int_{t-2z_2}^{\infty} t\, p_1(s,\underline{y}_1)\, p_2(t,\underline{y}_2)ds\, dt\Big]$$

$$= \frac{1}{4}\Big[\int_{-\infty}^{\infty} s\, p_1(s,\underline{y}_1)\, P_2(s+2z_2,\underline{y}_2)ds + \int_{-\infty}^{\infty} t\, p_2(t,\underline{y}_2)(1-P_1(t-2z_2,\underline{y}_1))dt\Big]$$

$$= F(z_2,\underline{y}) - E(z_2,\underline{y}).$$

The first and the second equalities follow from a change of variables and the third and the last equalities follow from definitions (3.5.15).

In a similar manner, we can show

$$\int_{-\infty}^{\infty} v\, g(v,\underline{y})dv = -\,E(Z_2,\underline{y}) - F(z_2,\underline{y}).$$

Therefore, we have from (3.5.10), (3.5.11) and (3.5.15),

$$\gamma_1 = z_1 - H(z_2, \underline{y}) \tag{3.5.16}$$

and $$\gamma_2 = z_2 - K(z_2, \underline{y}) \tag{3.5.17}$$

whenever $G(z_2, \underline{y}) > 0$, and $\gamma_1 = z_1$ and $\gamma_2 = z_2$ otherwise.

We need some lemmas before stating the main theorem. These lemmas are generalizations of Lemmas 3.1 and 3.2 and Theorem 3.3 of Blumenthal and Cohen (1968b) in the sense that $f(x-\theta_1)$ and $f(x-\theta_2)$ these, are replaced by $f_1(x-\theta_1)$ and $f_2(x-\theta_2)$. The proofs of Blumenthal and Cohen require some simple changes. We prove these lemmas for the sake of completeness.

<u>Lemma 3.5.2</u>: Let γ_1 and γ_2 be the components determining the Pitman estimator as defined in (3.5.2). Then for each $(z_1, \underline{y})$

$$\lim_{z_2 \to \infty} (\gamma_1 - z_1) = 0, \tag{3.5.18}$$

and $$\lim_{z_2 \to \infty} (\gamma_2 - z_2) = 0. \tag{3.5.19}$$

Whenever $G(z_2, \underline{y}) > 0$, $\gamma_2 - z_2$ is a nonincreasing function of z_2, and

$$|H(z_2, \underline{y})| \le -K(z_2, \underline{y}). \tag{3.5.20}$$

Also, $$\lim_{\mu_1 \to \infty} \lim_{\mu_2 \to \infty} R(\mu_1, \mu_2, \gamma_1, \gamma_2) = R \tag{3.5.21}$$

<u>Proof</u>: We have

$$\gamma_1 - z_1 = -\frac{\int_{-\infty}^{z_2} \int_{-\infty}^{\infty} u\, p_1(u-v, \underline{y}_1)\, p_2(u+v, \underline{y}_2)\, du\, dv}{\int_{-\infty}^{z_2} \int_{-\infty}^{\infty} p_1(u-v, \underline{y}_1)\, p_2(u+v, \underline{y}_2)\, du\, dv}$$

and

$$\gamma_2 - z_2 = - \frac{\int_{-\infty}^{z_2} v\, g(v,\underline{y})\, dv}{\int_{-\infty}^{z_2} g(v,\underline{y})\, dv} .$$

Using the fact that $\int_{-\infty}^{\infty} x\, p_i(x,\underline{y}_i)dx = 0$, $i = 1,2$, the statements (3.5.18) and (3.5.19) can be proved. To show that $\gamma_2 - z_2$ is non-increasing in z_2, we consider

$$\frac{d}{dz_2}(\gamma_2 - z_2) = \frac{- z_2\, g(z_2,\underline{y}) \int_{-\infty}^{z_2} g(v,\underline{y})dv + g(z_2,\underline{y}) \int_{-\infty}^{z_2} v\, g(v,\underline{y})dv}{G^2(z_2,\underline{y})}$$

$$= \frac{g(z_2,\underline{y})}{G^2(z_2,\underline{y})} \left[\int_{-\infty}^{z_2} (v-z_2)\; g(v,\underline{y})dv \right]$$

which is negative as g and G are positive functions and so $\gamma_2 - z_2$ is nonincreasing function of z_2.

The inequality (3.5.20), by definition (3.5.15), is equivalent to $F(z_2,\underline{y}) \geq 0$ and $E(z_2,\underline{y}) \geq 0$. Now,

$$F(z_2,\underline{y}) = \frac{1}{4} \int_{-\infty}^{\infty} x\, p_1(x,\underline{y}_1)\; P_2(x+2z_2,\underline{y}_2)dx$$

$$= \frac{1}{4} \int_{-\infty}^{\infty} \int_{-\infty}^{x+2z_2} x\, p_1(x,\underline{y}_1)\; p_2(t,\underline{y}_2)dt\, dx \qquad \text{(by (3.5.15))}$$

$$= \frac{1}{4} \int_{-\infty}^{\infty} \{ \int_{t-2z_2}^{\infty} x\, p_1(x,\underline{y}_1)dx \}\; p_2(t,\underline{y}_2)dt$$

by an interchange in the order of integration. For each t, the inner integral is always positive as $\int_{-\infty}^{\infty} x\, p_1(x,\underline{y}_1)dx = 0$. Therefore, $F(z_2,\underline{y}) \geq 0$.

Similarly, $E(z_2,\underline{y}) \geq 0$. This proves (3.5.20).

To prove (3.5.21), consider $R_1(\mu_1,\mu_2,\gamma_1)$ of (3.5.4), which by (3.5.16) is

$$R_1(\mu_1,\mu_2,\gamma_1) = \int_{\mathcal{Y}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1-H(z_2,\underline{y})-\mu_1)^2\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)dz_1\, dz_2\, \nu(d\underline{y})$$

$$\leq \int_{\mathcal{Y}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1-\mu_1)^2\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)dz_1\, dz_2\, \nu(d\underline{y})$$

$$+ \int_{\mathcal{Y}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H^2(z_2,\underline{y})\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)dz_1\, dz_2\, \nu(d\underline{y})$$

$$+ 2\int_{\mathcal{Y}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (z_1-\mu_1)\, H(z_2,\underline{y})\, p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1)$$
$$p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)dz_1\, dz_2\, \nu(d\underline{y})$$

$$= \frac{R}{8} + I_1(\mu_1,\mu_2) + I_2(\mu_1,\mu_2), \quad \text{say,} \tag{3.5.22}$$

where I_1 and I_2 denote the second and the third integrals respectively.

We proved in (3.5.20) that $|H(z_2,\underline{y})| \leq -K(z_2,\underline{y})$ for all $(z_2,\underline{y})$ such that $G(z_2,\underline{y}) > 0$. This implies that $H^2(z_2,\underline{y}) \leq K^2(z_2,\underline{y})$ for all $(z_2,\underline{y})$ such that $G(z_2,\underline{y}) > 0$. We also proved that $-K(z_2,\underline{y})$ decreases to zero as z_2 increases to ∞. A change of variables inside the integrals I_1 and I_2 and an application of the monotone convergence theorem, then, proves that

$$\lim_{\mu_2\to\infty} I_1(\mu_1,\mu_2) = 0 \quad \text{and}$$
$$\lim_{\mu_2\to\infty} I_2(\mu_1,\mu_2) = 0. \tag{3.5.23}$$

Therefore, we have $\lim_{\mu_2\to\infty} R_1(\mu_1,\mu_2,\gamma_1) = \frac{R}{8}$. (3.5.24)

Relations (3.5.3), (3.5.14) and (3.5.24) now prove (3.5.21).

Lemma 3.5.3: Suppose for each $\underline{y}$, there are real numbers $b(\underline{y})$ and $c(\underline{y})$ satisfying

$$b(\underline{y}) < c(\underline{y}) \quad \text{and} \quad \int_{b(\underline{y})}^{c(\underline{y})} g(z_2,\underline{y})dz_2 = \tfrac{1}{2} . \tag{3.5.25}$$

Also, let $\hat{E}(z_2,\underline{y})$ and $\hat{F}(z_2,\underline{y})$ be monotone functions of z_2 for each $\underline{y}$ such that $G(z_2,\underline{y}) > 0$. Denote the risk of the Pitman estimator $\underline{\delta}_p$ by $R(\mu_1,\mu_2,\gamma_1,\gamma_2)$ where μ_1, μ_2, γ_1 and γ_2 are given by (3.5.2). Then

$$R(\mu_1,\mu_2,\gamma_1,\gamma_2) \leq R \quad \text{if } \mu_2 \geq 0 . \tag{3.5.26}$$

Proof: Define,

$$\begin{aligned} R(\mu_1,\mu_2,\gamma_1,\gamma_2,\underline{y}) = 4 \int_{b(\underline{y})}^{c(\underline{y})} \int_{-\infty}^{\infty} & [(\gamma_1(z_2+\mu_2,z_1,\underline{y})-\mu_1)^2 \\ & + (\gamma_2(z_2+\mu_2,z_1,\underline{y})-\mu_2)^2] \\ & p_1((z_1-\mu_1)-z_2,\underline{y}_1)\, p_2((z_1-\mu_1)+z_2,\underline{y}_2)dz_1\, dz_2 \end{aligned} \tag{3.5.27}$$

and

$$R(\underline{y}) = 4 \int_{b(\underline{y})}^{c(\underline{y})} \int_{-\infty}^{\infty} (z_1^2+z_2^2)\, p_1(z_1-z_2,\underline{y}_1)\, p_2(z_1+z_2,\underline{y}_2)dz_1\, dz_2 \tag{3.5.28}$$

From the expressions (3.5.10) and (3.5.11) for γ_1 and γ_2 and the conditions $\int_{-\infty}^{\infty} x\, p_i(x,\underline{y}_i)dx = 0$, $i = 1,2$ and $\int_{b(\underline{y})}^{c(\underline{y})} g(z_2,\underline{y})dz_2 = 1$, we have $\gamma_1 = z_1$ and $\gamma_2 = z_2$ whenever $z_2 > c(\underline{y})$. Thus

$$R(\mu_1,\mu_2,\gamma_1,\gamma_2,\underline{y}) = R(\underline{y}) \quad \text{if } \mu_2 > c(\underline{y})-b(\underline{y}) \tag{3.5.29}$$

Next, we define

$$r(d) = \int_0^{\infty} (R(\underline{y}) - R(\mu_1,\mu_2+d,\gamma_1,\gamma_2,\underline{y}))d\mu_2. \tag{3.5.30}$$

We will show that $\Gamma(d)$ is decreasing in d. Let $0 < d_1 < d_2$, then

$$\Gamma(d_1)-\Gamma(d_2) = \int_0^\infty [R(\underline{y})-R(\mu_1,\mu_2+d_1,\gamma_1,\gamma_2,\underline{y})]\,d\mu_2$$
$$- \int_0^\infty [R(\underline{y})-R(\mu_1,\mu_2+d_2,\gamma_1,\gamma_2,\underline{y})]\,d\mu_2 .$$

The two integrals are finite as the integrands vanish for $\mu_2 > c(\underline{y})-b(\underline{y})-d_1$ and $\mu_2 > c(\underline{y})-b(\underline{y})-d_2$ respectively, and so we can combine them to get

$$\Gamma(d_1)-\Gamma(d_2) = \int_0^\infty [R(\mu_1,\mu_2+d_2,\gamma_1,\gamma_2,\underline{y}) - R(\mu_1,\mu_2+d_1,\gamma_1,\gamma_2,\underline{y})]\,d\mu_2$$

From relations (3.5.27), (3.5.16) and (3.5.17) we get

$$\begin{aligned}\Gamma(d_1)-\Gamma(d_2) = 4\int_0^\infty\int_{-\infty}^\infty\int_{-\infty}^\infty & [(z_1-H(z_2+\mu_2+d_2,\underline{y})-\mu_1)^2 \\ & + (z_2+\mu_2+d_2-K(z_2+\mu_2+d_2,\underline{y})-\mu_2-d_2)^2 \\ & - (z_1-H(z_2+\mu_2+d_1,\underline{y})-\mu_1)^2 \\ & - (z_2+\mu_2+d_1-K(z_2+\mu_2+d_1,\underline{y})-\mu_2-d_1)^2] \\ & p_1((z_1-\mu_1)-(z_2-\mu_2),\underline{y}_1) \\ & p_2((z_1-\mu_1)+(z_2-\mu_2),\underline{y}_2)\,dz_1\,dz_2\,d\mu_2 .\end{aligned}$$

A change of variables and some simplification yields,

$$\begin{aligned}\Gamma(d_1)-\Gamma(d_2) = 4\int_0^\infty\int_{-\infty}^\infty & \{[H^2(z_2+d_2,\underline{y})-H^2(z_2+d_1,\underline{y})]\ g(z_2-\mu_2,\underline{y}) \\ & + [K^2(z_2+d_2,\underline{y})-K^2(z_2+d_1,\underline{y})]\ g(z_2-\mu_2,\underline{y}) \\ & - 2[H(z_2+d_2,\underline{y})-H(z_2+d_1,\underline{y})]\ T(z_2-\mu_2,\underline{y}) \\ & - 2[K(z_2+d_2,\underline{y})-K(z_2+d_1,\underline{y})](z_2-\mu_2)\ T(z_2-\mu_2,\underline{y})\}\,dz_2\,d\mu_2\end{aligned}$$

The above double integral is absolutely convergent due to (3.5.29) and the fact that $g(z_2,\underline{y})$ vanishes for values of z_2 outside the

interval $(b(\underline{y}),c(\underline{y}))$. Therefore, we can interchange the order of integration. Making a transformation $z_2-\eta \to \eta$, and using (3.5.15), (3.5.16) and (3.5.17), we get

$$\begin{aligned}\Gamma(d_1)-\Gamma(d_2) = 4 \int_{-\infty}^{\infty} \{&[H^2(z_2+d_2,\underline{y})-H^2(z_2+d_1,\underline{y})]\ G(z_2,\underline{y}) \\ &- 2[H(z_2+d_2,\underline{y})-H(z_2+d_1,\underline{y})]\ H(z_2,\underline{y})\ G(z_2,\underline{y}) \\ &+ [K^2(z_2+d_2,\underline{y})-K^2(z_2+d_1,\underline{y})]\ G(z_2,\underline{y}) \\ &- 2[K(z_2+d_2,\underline{y})-K(z_2+d_1,\underline{y})]\ K(z_2,\underline{y})\ G(z_2,\underline{y})\}\ dz_2\end{aligned} \tag{3.5.31}$$

Once again using definitions for $H(z_2,\underline{y})$ and $K(z_2,\underline{y})$ and simplifying (3.5.31), we get

$$\begin{aligned}\Gamma(d_1)-\Gamma(d_2) = 8 \int_{-\infty}^{\infty} G(z_2,\underline{y})\{&[\hat{E}(z_2+d_2,\underline{y})-\hat{E}(z_2+d_1,\underline{y})]^2 \\ &+ [\hat{F}(z_2+d_2,\underline{y})-\hat{F}(z_2+d_1,\underline{y})]^2 \\ &+ 2[\hat{E}(z_2+d_2,\underline{y})-\hat{E}(z_2+d_1,\underline{y})][\hat{E}(z_2+d_1,\underline{y})-\hat{E}(z_2,\underline{y})] \\ &+ 2[\hat{F}(z_2+d_2,\underline{y})-\hat{F}(z_2+d_1,\underline{y})][\hat{F}(z_2+d_1,\underline{y})-\hat{F}(z_2,\underline{y})]\}\ dz_2\end{aligned}$$

Monotonicity of the functions $\hat{E}(z_2,\underline{y})$ and $\hat{F}(z_2,\underline{y})$ now proves that $\Gamma(d)$ is a nonincreasing function of d, that is, $\Gamma(d_1) \geq \Gamma(d_2)$ for $0 \leq d_1 < d_2$, which is equivalent to

$$\int_{d_1}^{d_2} (R(\underline{y}) - R(\mu_1,\mu_2,\gamma_1,\gamma_2,\underline{y}))\ d\mu_2 \geq 0 . \tag{3.5.32}$$

Since $R(\mu_1,\mu_2,\gamma_1,\gamma_2,\underline{y})$ is a continuous function of μ_2, (3.5.32) yields

$$R(\underline{y}) \geq R(\mu_1,\mu_2,\gamma_1,\gamma_2,\underline{y}) \quad \text{for all } \mu_2 \geq 0 .$$

Integrating on both the sides with respect to the measure ν, we get

$$R(\theta, \underline{\delta}_p) \leq R \quad \text{for all } \mu_2 \geq 0 .$$

Lemma 3.5.4: Let the density $p_i(x, \underline{y}_i)$, $i = 1, 2$, defined in (3.2.4) have finite variance and the functions $\hat{E}(z_2, \underline{y})$ and $\hat{F}(z_2, \underline{y})$ be monotone in z_2 for each $\underline{y}$ such that $G(z_2, \underline{y}) > 0$. Then the Pitman estimator $\underline{\delta}_p$ is minimax.

Proof: In view of Corollary 3.2.2, it suffices to show that

$$R(\theta, \underline{\delta}_p) \leq R \quad \text{for all } \theta_1 \leq \theta_2 .$$

Define density $p_n(z_o, z_1, \underline{y})$ for each $n \geq 1$ and for all $\underline{y}$ by

$$p_n(z_1, z_2, \underline{y}) = p_{1n}(z_1 - z_2, \underline{y}_1)\, p_{2n}(z_1 + z_2, \underline{y}_2),$$

where
$$p_{1n}(z_1 - z_2, \underline{y}_1)\, p_{2n}(z_1 + z_2, \underline{y}_2) = \begin{cases} p_1(z_1 - z_2, \underline{y}_1)\, p_2(z_1 + z_2, \underline{y}_2) & \text{for } |z_2| \leq n \\ 0 & \text{otherwise.} \end{cases}$$

Hereafter, we proceed as in Farrell (1964, pp. 985-986) with some changes necessitated by the fact that Farrell considers estimation of a location parameter θ of a density $f(x-\theta, y)$, $\theta > 0$, whereas we are estimating parameters θ_1 and θ_2 of densities $f_1(x_1 - \theta_1, \underline{y}_1)$ and $f_2(x_2 - \theta_2, \underline{y}_2)$, $\theta_1 \leq \theta_2$.

Theorem 3.5.5: Let the densities $p_i(x, \underline{y}_i)$ of (3.2.4) satisfy the following conditions:

(i) $q_1(x, \underline{y}_1) = \dfrac{p_1(x, \underline{y}_1)}{1 - P_1(x, \underline{y}_1)}$ is an increasing function of x for each $\underline{y}_1$.

(ii) $r_2(x, \underline{y}_2) = \dfrac{p_2(x, \underline{y}_2)}{P_2(x, \underline{y}_2)}$ is a decreasing function of x for each $\underline{y}_2$.

Then $\underline{\delta}_p$ is minimax.

Proof: Following Lemma 3.5.4, it suffices to show that the conditions (i) and (ii) imply the monotonicity of the functions $\hat{E}(z_2,\underline{y})$ and $\hat{F}(z_2,\underline{y})$ in z_2 for each fixed $\underline{y}$, for which $G(z_2,\underline{y}) > 0$. By definition, $\hat{F}(z_2^{(1)},\underline{y}) - \hat{F}(z_2^{(2)},\underline{y}) \geq 0$ is equivalent to

$$\frac{\int_{-\infty}^{\infty} x\, p_1(x,\underline{y}_1) P_2(x+2z_2^{(1)},\underline{y}_2)dx}{\int_{-\infty}^{\infty} p_1(x,\underline{y}_1)\ P_2(x+2z_2^{(1)},\underline{y}_2)dx} - \frac{\int_{-\infty}^{\infty} x\, p_1(x,\underline{y}_1) P_2(x+2z_2^{(2)},\underline{y}_2)dx}{\int_{-\infty}^{\infty} p_1(x,\underline{y}_1)\ P_2(x+2z_2^{(2)},\underline{y}_2)dx} \geq 0,$$

or

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x[P_2(x+2z_2^{(1)},\underline{y}_2)P_2(u+2z_2^{(2)},\underline{y}_2)$$
$$- P_2(x+2z_2^{(2)},\underline{y}_2)P_2(u+2z_2^{(1)},\underline{y}_2)]\ p_1(x,\underline{y}_1)p_2(x,\underline{y}_1)du\ dx \geq 0,$$

which by a break up of the inner integral in u in intervals $(-\infty,x)$ and (x,∞) and then by an interchange of order of integration is equivalent to

$$\int_{-\infty}^{\infty}\int_{-\infty}^{x} (x-u)[P_2(x+2z_2^{(1)},\underline{y}_2)P_2(u+2z_2^{(2)},\underline{y}_2)$$
$$- P_2(x+2z_2^{(2)},\underline{y}_2)P_2(u+2z_2^{(1)},\underline{y}_2)]\ p_1(x,\underline{y}_1)\ p_2(x,\underline{y}_2)du\ dx. \qquad (3.5.33)$$

Condition (ii) implies that $\log P_2(\cdot,\underline{y}_2)$ is concave for each $\underline{y}_2$ and so for $u < x$ and $z_2^{(1)} < z_2^{(2)}$ the integrand in (3.5.33) is nonnegative. Thus $\hat{F}(z_2,\underline{y})$ is nonincreasing for each $\underline{y}$. In a similar way, $\hat{E}(z_2,\underline{y})$ can be shown to be monotone for each $\underline{y}$.

This proves the minimaxity of $\underline{\delta}_p$.

As an application of Theorem 3.5.5, we have the following result.

Corollary 3.5.6: Let the density of $f_i(x-\theta_i)$, $i = 1,2$, have any of the following forms:

(i) $f_i(x) = \frac{1}{\sigma_i}\varphi(\frac{x}{\sigma_i}),\ \sigma_i > 0,\ -\infty < x < \infty$

(ii) $f_i(x) = \frac{1}{2a_i},\quad -a_i < x < a_i,$

(iii) $f_i(x) = \beta_i^{\alpha+1} x^{\alpha} \exp(-\beta_i x)/\Gamma(\alpha+1),\ x > 0,\ \alpha > 0,\ \beta_i > 0.$

Then the Pitman estimator is minimax.

Remark 3.5.1: Theorem 3.5.1 is quite restrictive in applications. Blumenthal and Cohen (1968b) state that the condition for Theorem 3.5.1 is satisfied by the uniform distribution considered in Corollary 3.5.6(ii) with $a_1 = a_2 = 1$. This is not true as shown below. The density $p(x,\underline{y})$ is (see Blumenthal and Cohen (1968a, p. 510)),

$$p(x,\underline{y}_i) = \frac{1}{2-R_n(\underline{y}_i)},\quad -1 + \frac{R_n(\underline{y}_i)}{2} \le x \le 1 - \frac{R_n(\underline{y}_i)}{2},$$

where

$$R_n(\underline{y}_i) = \max(0, y_{i1},\ldots,y_{i,n-1}) - \min(0, y_{i1},\ldots,y_{i,n-1}).$$

Now

$$\int_{-\infty}^{\infty} u\, p(u-v,\underline{y}_1)\, p(u+v,\underline{y}_2)\,du$$

$$= \int_{-\infty}^{\infty} \frac{u}{(2-R_n(\underline{y}_1))(2-R_n(\underline{y}_2))} I_{A(\underline{y}_1)}(u-v)\, I_{A(\underline{y}_2)}(u+v)\,du, \tag{3.5.34}$$

where $I_B(x)$ denotes the indicator function of the set B and

$$A(\underline{y}) = [-1 + \frac{R_n(\underline{y})}{2},\ 1 - \frac{R_n(\underline{y})}{2}].$$

It is obvious that the integral in expression (3.5.34) is not necessarily zero.

The minimaxity claim, however, is true as mentioned in Corollary 3.5.6.

CHAPTER - 4

Simultaneous and Componentwise Estimation of the Ordered Means of k Normal Populations

4.1 Introduction

In Chapter 3, the Pitman estimator $\underline{\delta}_p$ of the mean $\underline{\theta} = (\theta_1, \theta_2)$ of two normal populations with equal or unequal variances was seen to be minimax. Also, in general, when the two populations have location densities $f_1(x-\theta_1)$ and $f_2(x-\theta_2)$, sufficient conditions for the minimaxity of the Pitman estimator were derived. In this chapter, we consider this question for more than two normal populations. The problem of estimating components of $\underline{\theta}$ is also considered.

In Section 4.2, we consider simultaneous estimation of the means $\theta_1, \ldots, \theta_k$, $\theta_1 \leq \theta_2 \leq \ldots \leq \theta_k$; of k independent normal random variables with common known variance. The loss function is taken to be the squared error. The Pitman estimator $\underline{\delta}_p$ of (3.1.3) is shown to be minimax for any k. The proof uses an identity originally given in Stein (1973) and later generalized by Hudson (1978) and Stein (1981). A class of minimax estimators is also given. However, these estimators are not generalized Bayes and so inadmissible.

Inadmissibility of $\underline{\delta}_p$ follows from the results of Brown (1971, pp. 898-899) for $k \geq 3$. However, Brown does not give any dominating estimators. Following the heuristic approach of Brown (1979) to prove inadmissibility we obtain an estimator $\underline{\delta}'(\underline{X})$, which we feel, may possibly improve $\underline{\delta}_p$. We also use the Brewster-Zidek (1974) technique to prove that $\underline{\delta}_p$ is admissible among its multiples and to obtain a sufficient condition for its inadmissibility in a certain class of estimators.

In Section 4.3, the problem of estimating a component θ_i of $\underline{\theta}$ is taken up. When $k = 2$ and the two normal populations have unequal variances, Cohen and Sackrowitz (1970) proved that the component δ_{p2} of $\underline{\delta}_p$ for estimating θ_2 is minimax with respect to squared error loss. A symmetry argument proves minimaxity of δ_{p1} for θ_1. However, a different situation is encountered when $k = 3$. We show that the components δ_{p1} and δ_{p3} of $\underline{\delta}_p$ for estimating θ_1 and θ_3 respectively, are not minimax. The minimaxity of δ_{p2} for θ_2 is an open question.

In the end, we consider estimation of the larger location parameter θ_2 when the underlying densities are $f(x-\theta_1)$ and $f(x-\theta_2)$, $\theta_1 \le \theta_2$. When f is symmetric about 0, we prove an inadmissibility result which generalizes Theorems 2.1 and 5.1 of Cohen and Sackrowitz (1970).

4.2 Simultaneous Estimation

Let $X_1, X_2, \ldots, X_k$ be independent normal random variables with means $\theta_1, \theta_2, \ldots, \theta_k$ respectively; $\theta_1 \le \theta_2 \le \cdots \le \theta_k$, and the common variance unity. We estimate $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ with squared error loss,

$$L(\underline{\theta}, \underline{a}) = |\underline{\theta} - \underline{a}|^2 .$$

Minimaxity of the Pitman Estimator

We show in this section that the Pitman estimator $\underline{\delta}_p$ is minimax. The form of $\underline{\delta}_p = (\delta_{p1}, \ldots, \delta_{pk})$ is

$$\delta_{pi}(\underline{x}) = \int_{\Omega_0} \theta_i \, p(\underline{x}, \underline{\theta}) d\underline{\theta} \Big/ \int_{\Omega_0} p(\underline{x}, \underline{\theta}) d\underline{\theta} , \quad i = 1, 2, \ldots, k, \tag{4.2.1}$$

where $p(\underline{x},\underline{\theta}) = \varphi(x_1-\theta_1)\ \varphi(x_2-\theta_2)\ \dots\ \varphi(x_k-\theta_k)$,

$$\Omega_o = \{\underline{\theta} = (\theta_1,\dots,\theta_k):\ \theta_1 \le \theta \le \dots \le \theta_k\}, \tag{4.2.2}$$

and $d\underline{\theta}$ denotes $d\theta_1 \dots d\theta_k$.

Define

$$\alpha_i(\underline{x}) = \int_{\Omega_o} (\theta_i - x_i)\ p(\underline{x},\underline{\theta})d\underline{\theta}\ , \quad i = 1,\dots,k \tag{4.2.3}$$

and $$D(\underline{x}) = \int_{\Omega_o} p(\underline{x},\underline{\theta})d\underline{\theta}\ . \tag{4.2.4}$$

Then $$\delta_{pi}(x) = x_i + \gamma_i(\underline{x})\ , \tag{4.2.5}$$

where $$\gamma_i(\underline{x}) = \frac{\alpha_i(\underline{x})}{D(\underline{x})}\ , \quad i = 1,\dots,k\ . \tag{4.2.6}$$

Notice that $$\alpha_i(\underline{x}) = \frac{\partial}{\partial x_i} D(\underline{x})\ . \tag{4.2.7}$$

Also define,

$$u_o = x_1,\quad u_1 = x_2-x_1,\quad u_2 = x_3-x_2,\dots,u_{k-1} = x_k-x_{k-1}\ ,$$

and $$\tag{4.2.8}$$

$$\mu_o = \theta_1,\quad \mu_1 = \theta_2-\theta_1,\quad \mu_2 = \theta_3-\theta_2,\dots,\mu_{k-1} = \theta_k-\theta_{k-1}\ .$$

The parameter μ_i is nonnegative for $i = 1,2,\dots,k-1$. In terms of notation (4.2.8), $D(\underline{x})$ can be written as

$$D^*(\underline{u}) = \int_0^\infty \dots \int_0^\infty \int_{-\infty}^\infty \varphi(u_o-\mu_o)\ \varphi(u_o+u_1-\overline{\mu_o+\mu_1})\ \dots$$

$$\varphi(u_o+u_1+\dots+u_{k-1}-\overline{\mu_o+\mu_1+\dots+\mu_{k-1}})\ d\mu_o\ d\mu_1\dots d\mu_{k-1}, \tag{4.2.9}$$

where $\underline{u} = (u_1,\dots,u_{k-1})$. Observe that D^* does not depend on u_o.

Transforming the variables $\mu_o,\mu_1,\dots,\mu_{k-1}$ to $u_o-\mu_o, u_1-\mu_1, \dots, u_{k-1}-\mu_{k-1}$ respectively inside the integral (4.2.9), we get

$$D^*(\underline{u}) = \int_{-\infty}^{u_{k-1}} \cdots \int_{-\infty}^{u_1} \int_{-\infty}^{\infty} \varphi(\mu_o)\, \varphi(\mu_o+\mu_1) \cdots \varphi(\mu_o+\mu_1+\cdots+\mu_{k-1})\, d\mu_o\, d\mu_1 \cdots d\mu_{k-1}\,. \tag{4.2.10}$$

As an obvious consequence of the expression (4.2.10), we get the following result.

Lemma 4.2.1: $D^*(\underline{u})$ is an increasing function of each of the variables $u_1, u_2, \ldots, u_{k-1}$.

Before proving the main theorem of this section we state an identity due to Stein (1981).

Lemma 4.2.2: Let $\underline{Y}$ be a k dimensional normal random vector with mean vector $\underline{\eta}$ and the covariance matrix the identity matrix. Also let $\underline{\nabla}$ be the vector differential operator of the first order partial derivatives with i^{th} co-ordinate $\nabla_i = \frac{\partial}{\partial Y_i}$. If $h: R^k \to R$ is an almost differentiable function with $E_{\underline{\eta}} |\underline{\nabla}\, h(\underline{Y})| < \infty$, then

$$E_{\underline{\eta}}(\underline{\nabla}\, h(\underline{Y})) = E_{\underline{\eta}}\{(\underline{Y}-\underline{\eta})\, h(\underline{Y})\}.$$

Theorem 4.2.3: The Pitman estimator $\underline{\delta}_p$ is minimax.

Proof: The risk function of $\underline{\delta}_p$ is

$$R(\underline{\theta}, \underline{\delta}_p) = E\, |\underline{\delta}_p - \underline{\theta}|^2$$

$$= \sum_{i=1}^{k} E(X_i + \gamma_i(\underline{X}) - \theta_i)^2$$

$$= k + \sum_{i=1}^{k} E\{\gamma_i^2(\underline{X}) + 2E(X_i-\theta_i)\, \gamma_i(\underline{X})\}. \tag{4.2.11}$$

From Lemma 4.2.2, we get

$$E(X_i-\theta_i)\, \gamma_i(\underline{X}) = E\left(\frac{\partial}{\partial X_i}\, \gamma_i(\underline{X})\right)$$

$$= E\left(-\gamma_i^2(\underline{X}) + \frac{1}{D(\underline{X})}\, \frac{\partial}{\partial X_i}\, \alpha_i(\underline{X})\right)$$

using relations (4.2.6) and (4.2.7).

Therefore, we have from (4.2.11)

$$R(\underline{\theta},\delta_p) = k + \sum_{i=1}^{k} E(X_i \gamma_i(\underline{X}) + \frac{1}{D(\underline{X})}\frac{\partial}{\partial X_i}\alpha_i(\underline{X}) - \theta_i \gamma_i(\underline{X})) . \qquad (4.2.12)$$

We show below that

$$\sum_{i=1}^{k} (x_i \alpha_i(\underline{x}) + \frac{\partial}{\partial x_i}\alpha_i(\underline{x})) = 0 \quad \text{for each } \underline{x} \in R^k . \qquad (4.2.13)$$

From (4.2.3),

$$\sum_{i=1}^{k} (x_i \alpha_i(\underline{x}) + \frac{\partial}{\partial x_i}\alpha_i(\underline{x}))$$

$$= \sum_{i=1}^{k} [\, x_i \int_{\Omega_0} (\theta_i - x_i)p(\underline{x},\theta)d\underline{\theta} + \frac{\partial}{\partial x_i}\int_{\Omega_0} (\theta_i - x_i)p(\underline{x},\underline{\theta})d\underline{\theta}\,]$$

$$= \sum_{i=1}^{k} [\, x_i \int_{\Omega_0} (\theta_i - x_i)p(\underline{x},\underline{\theta})d\underline{\theta} + \int_{\Omega_0} ((\theta_i - x_i)^2 - 1)p(\underline{x},\underline{\theta})d\underline{\theta}\,] .$$

Combining the two integrals above, we get

$$\sum_{i=1}^{k} (x_i \alpha_i(\underline{x}) + \frac{\partial}{\partial x_i}\alpha_i(\underline{x}))$$

$$= \sum_{i=1}^{k} [\, \int_{\Omega_0} (\theta_i(\theta_i - x_i) - 1)p(\underline{x},\underline{\theta})d\underline{\theta}\,] . \qquad (4.2.14)$$

Consider integration by parts of $\theta_i(\theta_i - x_i)$ with respect to θ_i, $i = 1,\ldots,k$ on the right hand side of (4.2.14) to get

$$\int_{\Omega_0} (\theta_1(\theta_1 - x_1) - 1)p(\underline{x},\underline{\theta})d\underline{\theta}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_k} \cdots \int_{-\infty}^{\theta_3} \theta_2\, \varphi(\theta_2 - x_1)\, \varphi(\theta_2 - x_2) \cdots \varphi(\theta_k - x_k)\, d\theta_2 \cdots d\theta_k ,$$

$$\int_{\Omega_0} (\theta_2(\theta_2-x_2)-1)\,p(\underline{x},\underline{\theta})\,d\underline{\theta}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_k}\cdots\int_{-\infty}^{\theta_3} (\theta_3\;\varphi(\theta_3-x_2)-\theta_1\;\varphi(\theta_1-x_2))\varphi(\theta_1-x_1)\varphi(\theta_3-x_3)\;\cdots$$

$$\varphi(\theta_k-x_k)\,d\theta_1\,d\theta_3\;\cdots\;d\theta_k\;,$$

$$\cdots \qquad\qquad \cdots \qquad\qquad \cdots \qquad\qquad \cdots$$

$$\int_{\Omega_0} (\theta_k(\theta_k-x_k)-1)\,p(\underline{x},\underline{\theta})\,d\theta$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\theta_{k-1}}\cdots\int_{-\infty}^{\theta_2} \theta_{k-1}\;\varphi(\theta_{k-1}-x_k)\varphi(\theta_1-x_1)\;\cdots\;\varphi(\theta_{k-1}-x_{k-1})\,d\theta_1\cdots d\theta_{k-1}\,.$$

Adding the above identities, we get from (4.2.14),

$$\sum_{i=1}^{k} (x_i\alpha_i(\underline{x}) + \frac{\partial}{\partial x_i}\,\alpha_i(\underline{x})) = 0 \quad \text{for all } x \in R^k\,.$$

Next, we prove

$$\sum_{i=1}^{k} \alpha_i(\underline{x}) = 0 \quad \text{for all } \underline{x} \in R^k\,. \tag{4.2.15}$$

By definition (4.2.3)

$$\sum_{i=1}^{k} \alpha_i(\underline{x}) = \sum_{i=1}^{k} \int_{\Omega_0} (\theta_i-x_i)\;p(\underline{x},\underline{\theta})\,d\underline{\theta}\,. \tag{4.2.16}$$

As in the proof of (4.2.13), we integrate $(\theta_i-x_i)p(\underline{x},\underline{\theta})$ with respect to θ_i, $i = 1,\ldots,k$, on the right hand side of (4.2.16) and obtain the following relations

$$\int_{\Omega_o} (\theta_1-x_1)\,p(\underline{x},\underline{\theta})\,d\underline{\theta}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_k}\cdots\int_{-\infty}^{\theta_3} \varphi(\theta_2-x_1)\ \varphi(\theta_2-x_2)\cdots\varphi(\theta_k-x_k)\,d\theta_1\cdots d\theta_k\ ,$$

$$\int_{\Omega_o} (\theta_2-x_2)\,p(\underline{x},\underline{\theta})\,d\underline{\theta}$$

$$= -\int_{-\infty}^{\infty}\int_{-\infty}^{\theta_k}\cdots\int_{-\infty}^{\theta_3} (\varphi(\theta_3-x_2)-\varphi(\theta_1-x_2))\ \varphi(\theta_1-x_1)\ \varphi(\theta_3-x_3)\cdots$$
$$\varphi(\theta_k-x_k)\,d\theta_1\ d\theta_3\cdots d\theta_k\ ,$$

... ...

$$\int_{\Omega_o} (\theta_k-x_k)\,p(\underline{x},\underline{\theta})\,d\underline{\theta}$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\theta_{k-1}}\cdots\int_{-\infty}^{\theta_2} \varphi(\theta_{k-1}-x_k)\ \varphi(\theta_1-x_1)\cdots\varphi(\theta_{k-1}-x_{k-1})\,d\theta_1\cdots d\theta_{k-1}\ .$$

Adding the above relations we get (4.2.15).

Using (4.2.15) alongwith the definitions (4.2.8), one can easily verify the following relations among $\alpha_i(\underline{x})$ and $\frac{\partial}{\partial u_i} D^*(\underline{u})$, $i = 1,\ldots,k$,

$$\begin{aligned}
\alpha_1(\underline{x}) &= -\frac{\partial}{\partial u_1} D^*(\underline{u})\ , \\
\alpha_2(\underline{x}) &= \frac{\partial}{\partial u_1} D^*(\underline{u}) - \frac{\partial}{\partial u_2} D^*(\underline{u})\ , \\
&\cdots \qquad \cdots \qquad \cdots \\
\alpha_{k-1}(\underline{x}) &= \frac{\partial}{\partial u_{k-2}} D^*(\underline{u}) - \frac{\partial}{\partial u_{k-1}} D^*(\underline{u})\ ,
\end{aligned} \tag{4.2.17}$$

and $\quad \alpha_k(\underline{x}) = \frac{\partial}{\partial u_{k-1}} D^*(\underline{u})\ .$

From (4.2.17),

$$\sum_{i=1}^{k} \theta_i \alpha_i(\underline{x}) = -\theta_1 \frac{\partial}{\partial u_1} D^*(\underline{u}) + \theta_2 \frac{\partial}{\partial u_1} D^*(\underline{u}) - \theta_2 \frac{\partial}{\partial u_2} D^*(\underline{u}) + \dots + \theta_k \frac{\partial}{\partial u_{k-1}} D^*(\underline{u})$$

$$= \sum_{i=1}^{k-1} \mu_i \frac{\partial}{\partial u_i} D^*(\underline{u}) . \qquad (4.2.18)$$

Substituting from (4.2.13) and (4.2.18) in (4.2.12),

$$R(\underline{\theta}, \underline{\delta}_p) = k - \sum_{i=1}^{k-1} \mu_i \, E\{\frac{1}{D(\underline{X})} \frac{\partial D^*(\underline{U})}{\partial U_i}\} . \qquad (4.2.19)$$

The fact that μ_i's are nonnegative over Ω_o and Lemma 4.2.1 prove that

$$R(\underline{\theta}, \underline{\delta}_p) \leq k \quad \text{for all } \underline{\theta} \in \Omega_o . \qquad (4.2.20)$$

Since the risk of $\underline{X} = (X_1, \dots, X_k)$ is k, we conclude from Corollary 3.2.2 that $\underline{\delta}_p$ is a minimax estimator of $\underline{\theta}$.

Remark 4.2.1: An alternative expression for $\underline{\delta}_p$, as given below, can be obtained making use of the relations (4.2.5) and (4.2.17).

$$\underline{\delta}_p(\underline{x}) = \begin{bmatrix} x_1 - \frac{\partial}{\partial u_1} D^*(\underline{u}) \\ x_2 + \frac{\partial}{\partial u_1} D^*(\underline{u}) - \frac{\partial}{\partial u_2} D^*(\underline{u}) \\ \dots \quad \dots \quad \dots \\ x_{k-1} + \frac{\partial}{\partial u_{k-2}} D^*(\underline{u}) - \frac{\partial}{\partial u_{k-1}} D^*(\underline{u}) \\ x_k + \frac{\partial}{\partial u_{k-1}} D^*(\underline{u}) \end{bmatrix} . \qquad (4.2.21)$$

The Mixed Estimators

In Chapter 3, we studied the mixed estimators for estimating ordered means of two independent normal random variables both when

they have equal variances or unequal variances. However, the mixed estimators can be introduced for any number k of normal populations. The result for k = 2 can be used in steps to construct the mixed estimator for k > 2. In general we have the following result.

<u>Theorem 4.2.4</u>: If $\underline{\delta}(\underline{X}) = (\delta_1(\underline{X}),\ldots,\delta_k(\underline{X}))$ is an estimator of $\underline{\theta}$ such that

$P_{\underline{\theta}'}(\delta_1(\underline{X}) \le \delta_2(\underline{X}) \le \cdots \le \delta_k(\underline{X})) < 1$ for some $\underline{\theta}' \in \Omega_o$,

a mixed estimator of $\underline{\delta}$ can be constructed, which improves $\underline{\delta}$.

<u>Proof</u>: For convenience we take k = 3. For k > 3 we have to repeat the argument. Consider the following subsets of R^3,

$$A_1 = \{\underline{x}:\ \delta_1(\underline{x}) \le \delta_2(\underline{x}) \le \delta_3(\underline{x})\},$$

$$A_2 = \{\underline{x}:\ \delta_1(\underline{x}) < \delta_3(\underline{x}) < \delta_2(\underline{x})\},$$

$$A_3 = \{\underline{x}:\ \delta_2(\underline{x}) < \delta_1(\underline{x}) < \delta_3(\underline{x})\},$$

$$A_4 = \{\underline{x}:\ \delta_2(\underline{x}) < \delta_3(\underline{x}) < \delta_1(\underline{x})\},$$

$$A_5 = \{\underline{x}:\ \delta_3(\underline{x}) < \delta_1(\underline{x}) < \delta_2(\underline{x})\},$$

and $$A_6 = \{\underline{x}:\ \delta_3(\underline{x}) < \delta_2(\underline{x}) < \delta_1(\underline{x})\}.$$

The sets A_i, $i = 1,\ldots,6$ are disjoint and $P_{\underline{\theta}}(\bigcup_{i=1}^{6} A_i) = 1$ for all $\underline{\theta} \in \Omega_o$. If $P_{\underline{\theta}'}(A_1) < 1$ for some $\underline{\theta}' \in \Omega_o$, we have $P_{\underline{\theta}'}(A_i) > 0$ for at least one $i = 2,\ldots,6$. Let, for example, $P_{\underline{\theta}'}(A_2) > 0$, then we can use the construction of Lemma 3.3.1 to get an estimator $(\delta_1,\delta_2^*,\delta_3^*)$ satisfying $\delta_1(\underline{x}) \le \delta_2^*(\underline{x}) \le \delta_3^*(\underline{x})$ over A_2. In a similar way, for other regions A_i, $i = 3,\ldots,6$, we can obtain estimators lying in the space $\{\underline{\theta}:\ \theta_1 \le \theta_2 \le \theta_3\}$. Finally, we get an estimator, say $\underline{\delta}' = (\delta_1',\delta_2',\delta_3')$ which satisfies $\delta_1'(\underline{x}) \le \delta_2'(\underline{x}) \le \delta_3'(\underline{x})$ with probability one. This $\underline{\delta}'$ improves $\underline{\delta}$ because of Remark 3.3.2.

Remark 4.2.2: Consider the mixed estimator of $\underline{X}$. Since $\underline{X}$ is minimax, the mixed estimator is also minimax.

Remark 4.2.3: The risk function of the mixed estimator is complicated even for k = 3 and comparison with $\underline{\delta}_p$ is difficult to make.

Remark 4.2.4: By an argument similar to that in Theorem 3.3.5, we can show that the mixed estimators are not generalized Bayes and so are inadmissible.

Some Alternative Estimators

Let $\underline{X}_1,\ldots,\underline{X}_n$ be a random sample from a p-variate normal distribution with mean vector $\underline{\theta}$ and the covariance matrix identity matrix. For estimating $\underline{\theta}$, let the loss function be the square of the distance between the parameter and the estimate. The usual estimator, the sample mean $\underline{\bar{X}}$ is the best translation equivariant estimator of $\underline{\theta}$. For p = 1 and p = 2, $\underline{\bar{X}}$ is admissible. However, Stein (1956) proved the inadmissibility of $\underline{\bar{X}}$ for $p \geq 3$, a surprising result in view of the fact that the components of $\underline{\bar{X}}$ are independent and the loss in estimating $\underline{\theta}$ is the sum of individual losses in estimating θ_i's. Brown (1966) considered location parameter densities and general loss functions. He proved under mild conditions that the best translation equivariant estimator is admissible for p = 1 and 2 and inadmissible for $p \geq 3$.

For the problem of estimating two ordered location parameters we know that the Pitman estimator $\underline{\delta}_p$ is admissible under some conditions of finiteness of moments when the loss function is squared error (Blumenthal and Cohen (1968b), Theorem 4.1). It would be interesting to see what happens when we have more than

two populations. The result of Brown (1971, pp. 898-899) proves that for the problem of estimating k ordered normal means the Pitman estimator $\underline{\delta}_p$ is inadmissible for $k \geq 3$, although he does not provide a better estimator. In this section, we consider the case $k = 3$. We show that $\underline{\delta}_p = \underline{X} + \gamma(\underline{X})$ is admissible among its multiples and derive a necessary and sufficient condition for $\underline{\delta}_p$ to be admissible among estimators of the form $\underline{X} + c\,\underline{\gamma}(\underline{X})$, c a constant. We also suggest some estimators which may possibly improve $\underline{\delta}_p$.

When $k = 3$, we have $\underline{X} = (X_1, X_2, X_3)$, $\underline{\theta} = (\theta_1, \theta_2, \theta_3)$; $p(\underline{x}, \underline{\theta})$ and Ω_o of (4.2.2) are given by

$$p(\underline{x},\underline{\theta}) = \varphi(x_1-\theta_1)\,\varphi(x_2-\theta_2)\,\varphi(x_3-\theta_3)\,, \tag{4.2.22}$$

and $\Omega_o = \{\underline{\theta} = (\theta_1,\theta_2,\theta_3) : \theta_1 \leq \theta_2 \leq \theta_3\}$.

Also $$D(\underline{x}) = \int_{\Omega_o} \varphi(x_1-\theta_1)\,\varphi(x_2-\theta_2)\,\varphi(x_3-\theta_3)\,d\theta_1\,d\theta_2\,d\theta_3\,. \tag{4.2.23}$$

Using the identities (3.3.14) and (3.3.15) and simplifying, we get the α_i's of (4.2.3) as

$$\alpha_1(\underline{x}) = -\,Q_1(\underline{x})\,, \tag{4.2.24}$$

$$\alpha_2(\underline{x}) = Q_1(\underline{x}) - Q_2(\underline{x})\,,$$

and $\alpha_3(\underline{x}) = Q_2(\underline{x})$,

where $$Q_1(\underline{x}) = \frac{1}{\sqrt{2}}\,\varphi\left(\frac{x_1-x_2}{\sqrt{2}}\right)\,\Phi\left(\frac{2x_3-x_1-x_2}{\sqrt{6}}\right)\,, \tag{4.2.25}$$

and $$Q_2(\underline{x}) = \frac{1}{\sqrt{2}}\,\varphi\left(\frac{x_2-x_3}{\sqrt{2}}\right)\,\Phi\left(\frac{x_2+x_3-2x_1}{\sqrt{6}}\right)\,.$$

Therefore,

$$\underline{\delta}_p(\underline{x}) = \underline{x} + \underline{\gamma}(\underline{x})\,, \tag{4.2.26}$$

where,

$$\underline{\gamma}(\underline{x}) = \begin{bmatrix} -Q_1(\underline{x})/D(\underline{x}) \\ (Q_1(\underline{x}) - Q_2(\underline{x}))/D(\underline{x}) \\ Q_2(\underline{x})/D(\underline{x}) \end{bmatrix} . \tag{4.2.27}$$

A useful technique for improving upon estimators is by Brewster and Zidek (1974). An application of the technique leads to the result that $\underline{\delta}_p$ is admissible among its multiples.

<u>Theorem 4.2.5</u>: Among the estimators $c\,\underline{\delta}_p$, c a constant the only admissible estimators for the squared error loss are given by $c\,\underline{\delta}_p$, where $0 \le c \le 1$.

For a proof of the above theorem we need the following lemma.

<u>Lemma 4.2.6</u>: Let $D(\underline{x})$, $Q_1(\underline{x})$ and $Q_2(\underline{x})$ be as defined in (4.2.23) and (4.2.25). Then

$$2E\left[\frac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})} + (X_2-X_1)\frac{Q_1(\underline{X})}{D(\underline{X})} + (X_3-X_2)\frac{Q_2(\underline{X})}{D(\underline{X})}\right]$$

$$= (\theta_2-\theta_1)\,E\,\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)\,E\,\frac{Q_2(\underline{X})}{D(\underline{X})} . \tag{4.2.28}$$

<u>Proof</u>: Define

$$\begin{aligned} P(\underline{x}) &= \frac{1}{\sqrt{2}}\,\varphi\left(\frac{x_1-x_2}{\sqrt{2}}\right)\frac{1}{\sqrt{6}}\,\varphi\left(\frac{2x_3-x_1-x_2}{\sqrt{6}}\right) \\ &= \frac{1}{\sqrt{2}}\,\varphi\left(\frac{x_2-x_3}{\sqrt{2}}\right)\frac{1}{\sqrt{6}}\,\varphi\left(\frac{x_2+x_3-2x_1}{\sqrt{6}}\right) . \end{aligned} \tag{4.2.29}$$

Then

$$\frac{\partial Q_1(\underline{x})}{\partial x_1} = \left(\frac{x_2-x_1}{2}\right) Q_1(\underline{x}) - P(\underline{x}) ,$$

$$\frac{\partial Q_1(\underline{x})}{\partial x_2} = (\frac{x_1-x_2}{2})\ Q_1(\underline{x}) - P(\underline{x}) ,$$

$$\frac{\partial Q_2(\underline{x})}{\partial x_2} = (\frac{x_3-x_2}{2})\ Q_2(\underline{x}) + P(\underline{x}) , \tag{4.2.30}$$

and $$\frac{\partial Q_2(\underline{x})}{\partial x_3} = (\frac{x_2-x_3}{2})\ Q_2(\underline{x}) + P(\underline{x}) .$$

Now we apply Lemma 4.2.2 to evaluate $\theta_i\ E\ \frac{Q_1(\underline{X})}{D(\underline{X})}$, $i = 1,2$ and $\theta_j\ E\ \frac{Q_2(\underline{X})}{D(\underline{X})}$, $j = 2,3$.

$$\theta_1\ E(\frac{Q_1(\underline{X})}{D(\underline{X})}) = E(X_1 \frac{Q_1(\underline{X})}{D(\underline{X})} - \frac{\partial}{\partial X_1} \frac{Q_1(\underline{X})}{D(\underline{X})})$$

$$= E[\frac{1}{D(\underline{X})} \{X_1Q_1(\underline{X}) - ((\frac{X_2-X_1}{2})\ Q_1(\underline{X})-P(\underline{X}))\} - \frac{Q_1^2(\underline{X})}{D^2(\underline{X})}] , \tag{4.2.31}$$

$$\theta_2\ E(\frac{Q_1(\underline{X})}{D(\underline{X})}) = E(X_2 \frac{Q_1(\underline{X})}{D(\underline{X})} - \frac{\partial}{\partial X_2} \frac{Q_1(\underline{X})}{D(\underline{X})})$$

$$= E[\frac{1}{D(\underline{X})} \{X_2Q_1(\underline{X})-((\frac{X_1-X_2}{2})\ Q_1(\underline{X}) - P(\underline{X}))\}$$

$$+ \frac{Q_1(\underline{X})\,(Q_1(\underline{X})-Q_2(\underline{X}))}{D^2(\underline{X})}] , \tag{4.2.32}$$

$$\theta_2\ E(\frac{Q_2(\underline{X})}{D(\underline{X})}) = E(X_2 \frac{Q_2(\underline{X})}{D(\underline{X})} - \frac{\partial}{\partial X_2} \frac{Q_2(\underline{X})}{D(\underline{X})})$$

$$= E[\frac{1}{D(\underline{X})} \{X_2Q_2(\underline{X}) - ((\frac{X_3-X_2}{2})\ Q_2(\underline{X}) + P(\underline{X}))\}$$

$$+ \frac{Q_2(\underline{X})\,(Q_1(\underline{X})-Q_2(\underline{X}))}{D^2(\underline{X})}] , \tag{4.2.33}$$

and

$$\theta_3\ E(\frac{Q_2(\underline{X})}{D(\underline{X})}) = E(X_3 \frac{Q_2(\underline{X})}{D(\underline{X})} - \frac{\partial}{\partial X_3} \frac{Q_2(\underline{X})}{D(\underline{X})})$$

$$= E[\frac{1}{D(\underline{X})} \{X_3Q_2(\underline{X}) - ((\frac{X_2-X_3}{2})Q_2(\underline{X}) + P(\underline{X}))\} + \frac{Q_2^2(\underline{X})}{D^2(\underline{X})}] . \tag{4.2.34}$$

Subtracting (4.2.31) from (4.2.32) and (4.2.33) from (4.2.34) and then adding the resulting identities we get (4.2.28), as was to be proved.

Proof of Theorem 4.2.5: The risk of $c\,\underline{\delta}_p$ is

$$R(\underline{\theta},\ c\,\underline{\delta}_p) = E_{\underline{\theta}}\,|c\,\underline{\delta}_p - \underline{\theta}\,|^2 = h(c,\underline{\theta}), \text{ say.}$$

Notice that for fixed $\underline{\theta}$ risk is a convex function of c. The value of c minimizing $h(c,\underline{\theta})$ for fixed $\underline{\theta}$ is a solution of $\frac{\partial}{\partial c}\,h(c,\underline{\theta}) = 0$ and is given by

$$\hat{c}(\underline{\theta}) = \frac{E_{\underline{\theta}}(\theta_1\,\delta_{p1} + \theta_2\,\delta_{p2} + \theta_3\,\delta_{p3})}{E_{\underline{\theta}}(\delta_{p1}^2 + \delta_{p2}^2 + \delta_{p3}^2)}$$

$$= \frac{\theta_1 E(X_1 - \frac{Q_1(\underline{X})}{D(\underline{X})}) + \theta_2 E(X_2 + \frac{Q_1(\underline{X})-Q_2(\underline{X})}{D(\underline{X})}) + \theta_3 E(X_3 + \frac{Q_2(\underline{X})}{D(\underline{X})})}{E(X_1 - \frac{Q_1(\underline{X})}{D(\underline{X})})^2 + E(X_2 + \frac{Q_1(\underline{X})-Q_2(\underline{X})}{D(\underline{X})})^2 + E(X_3 + \frac{Q_2(\underline{X})}{D(\underline{X})})^2}$$

$$= \frac{\sum_{i=1}^{3}\theta_i^2 + (\theta_2-\theta_1)\,E\,\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)\,E\,\frac{Q_2(\underline{X})}{D(\underline{X})}}{3 + \sum_{i=1}^{3}\theta_i^2 + 2E\left[\frac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})} + \frac{(X_2-X_1)Q_1(\underline{X}) + (X_3-X_2)Q_2(\underline{X})}{D(\underline{X})}\right]}$$

$$= \frac{\sum_{i=1}^{3}\theta_i^2 + (\theta_2-\theta_1)\,E\,\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)\,E\,\frac{Q_2(\underline{X})}{D(\underline{X})}}{3 + \sum_{i=1}^{3}\theta_i^2 + (\theta_2-\theta_1)\,E\,\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)\,E\,\frac{Q_2(\underline{X})}{D(\underline{X})}}\,.$$

The last ~~in~~equality follows from Lemma 4.2.6.

For $\theta_1 \leq \theta_2 \leq \theta_3$ we have $0 \leq \hat{c}(\underline{\theta}) < 1$ and $\hat{c}(\underline{\theta})$ approaches one as θ_1 approaches infinity along the line $\theta_1 = \theta_2 = \theta_3$.

Hence $\inf_{\underline{\theta} \in \Omega_o} \hat{c}(\underline{\theta}) = \hat{c}(\underline{0}) = 0$ and $\sup_{\underline{\theta} \in \Omega_o} \hat{c}(\underline{\theta}) = 1.$

Any value of c in the interval $[0,1)$ is a $\hat{c}(\underline{\theta})$ and so minimizes the risk at some $\underline{\theta}$. Therefore, such a $c\underline{\delta}_p$ is admissible among multiples of $\underline{\delta}_p$. Also $c\underline{\delta}_p$, $c < 0$ is improved by $\underline{0}$ and $c\underline{\delta}_p$, $c > 1$ is improved by $\underline{\delta}_p$.

Next, we prove that $\underline{\delta}_p$ is admissible among $c\underline{\delta}_p$.

Let $c_o\,\underline{\delta}_p$, $0 \le c_o < 1$ satisfy

$$R(\underline{\theta}, c_o\,\underline{\delta}_p) \le R(\underline{\theta}, \underline{\delta}_p) \quad \text{for all } \theta_1 \le \theta_2 \le \theta_3 . \tag{4.2.35}$$

Consider $R(\underline{\theta}, c\underline{\delta}_p)$

$$= E_{\underline{\theta}} \mid c\underline{\delta}_p - \underline{\theta} \mid^2$$

$$= E(cX_1 - \frac{cQ_1(\underline{X})}{D(\underline{X})} - \theta_1)^2 + E(cX_2 + \frac{c(Q_1(\underline{X}) - Q_2(\underline{X}))}{D(\underline{X})} - \theta_2)^2 + E(cX_3 + \frac{cQ_2(\underline{X})}{D(\underline{X})} - \theta_3)^2$$

$$= E(cX_1-\theta_1)^2 + E(cX_2-\theta_2)^2 + E(cX_3-\theta_3)^2$$
$$+ 2E[-(cX_1-\theta_1)\frac{cQ_1(\underline{X})}{D(\underline{X})} + (cX_2-\theta_2)\frac{c(Q_1(\underline{X})-Q_2(\underline{X}))}{D(\underline{X})} + (cX_3-\theta_3)\frac{cQ_2(\underline{X})}{D(\underline{X})}]$$
$$+ 2c^2 E(\frac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})})$$

$$= 3c^2 + (c-1)^2 \sum_{i=1}^{3} \theta_i^2 - 2c[(\theta_2-\theta_1)E\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)E\frac{Q_2(\underline{X})}{D(\underline{X})}]$$
$$+ 2c^2 E[(X_2-X_1)\frac{Q_1(\underline{X})}{D(\underline{X})} + (X_3-X_2)\frac{Q_2(\underline{X})}{D(\underline{X})} + \frac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})}]$$

$$= 3c^2 + (c-1)^2 \sum_{i=1}^{3} \theta_i^2 + (c^2-2c)\,[(\theta_2-\theta_1)E\frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)E\frac{Q_2(\underline{X})}{D(\underline{X})}] \tag{4.2.36}$$

by Lemma 4.2.6.

Comparing (4.2.21) with (4.2.26) and (4.2.27), we have

$$\frac{\partial}{\partial u_1} D^*(\underline{u}) = Q_1(\underline{x}) \quad \text{and} \quad \frac{\partial}{\partial u_2} D^*(\underline{u}) = Q_2(\underline{x}),$$

and so from (4.2.19),

$$R(\underline{\theta}, \underline{\delta}_p) = 3 - (\theta_2-\theta_1) E \frac{Q_1(\underline{X})}{D(\underline{X})} - (\theta_3-\theta_2) E \frac{Q_2(\underline{X})}{D(\underline{X})} . \tag{4.2.37}$$

Substituting from (4.2.36) and (4.2.37) in (4.2.35), we get

$$3c_o^2 + (c_o-1)^2 \sum_{i=1}^{3} \theta_i^2 + (c_o^2-2c_o) \left[(\theta_2-\theta_1) E \frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2) E \frac{Q_2(\underline{X})}{D(\underline{X})} \right]$$

$$\leq 3 - \left[(\theta_2-\theta_1) E \frac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2) E \frac{Q_2(\underline{X})}{D(\underline{X})} \right] \quad \text{for all } \theta_1 \leq \theta_2 \leq \theta_3 .$$

When $\theta_1 = \theta_2 = \theta_3 = \theta$, say, the above inequality takes the form

$$c_o^2 + (c_o-1)^2 \theta^2 \leq 1 . \tag{4.2.38}$$

The left hand side of (4.2.38) approaches infinity as $\theta \to \infty$ whereas the right hand side is simply one, thus giving a contradiction. Hence, $\underline{\delta}_p$ is admissible among estimators of the form $c\underline{\delta}_p$.

Remark 4.2.5: On any compact subset Ω_o^* of Ω_o we can improve upon $\underline{\delta}_p$ by $c^*\underline{\delta}_p$, where $c^* = \sup_{\underline{\theta} \in \Omega_o^*} \hat{c}(\underline{\theta}) < 1$. This is so because $R(\underline{\theta}, c\underline{\delta}_p)$ is a convex function of c.

We employ the Brewster-Zidek technique on estimators of the form

$$\underline{d}_s(\underline{X}) = \underline{X} + s\,\underline{\gamma}(\underline{X}) ,$$

and obtain a sufficient condition for inadmissibility of $\underline{\delta}_p$.

The risk function of $\underline{d}_s$ is

$$R(\underline{\theta},\underline{d}_s) = E\,|\underline{X} + s\,\underline{\gamma}(\underline{X}) - \underline{\theta}|^2$$

$$= E|\underline{X}-\underline{\theta}|^2 + 2s\,E[(X_1-\theta_1)\gamma_1(\underline{X}) + (X_2-\theta_2)\gamma_2(\underline{X}) + (X_2-\theta_3)\gamma_3(\underline{X})$$

$$+ s^2\,E|\underline{\gamma}(\underline{X})|^2 ,$$

which for fixed $\underline{\theta}$ is a convex function of s. The value of s minimizing $R(\underline{\theta},\underline{d}_s)$ for each $\underline{\theta}$ is

$$\hat{s}(\underline{\theta}) = \frac{E[(\theta_1-X_1)\gamma_1(\underline{X}) + (\theta_2-X_2)\gamma_2(\underline{X}) + (\theta_3-X_3)\gamma_3(\underline{X})]}{2E\left[\dfrac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})}\right]}$$

$$= \frac{1}{2}\left[1 + \frac{1}{2}\left\{\frac{\left((\theta_2-\theta_1)E\dfrac{Q_1(\underline{X})}{D(\underline{X})} + (\theta_3-\theta_2)E\dfrac{Q_2(\underline{X})}{D(\underline{X})}\right)}{E\left(\dfrac{Q_1^2(\underline{X}) + Q_2^2(\underline{X}) - Q_1(\underline{X})Q_2(\underline{X})}{D^2(\underline{X})}\right)}\right\}\right]$$

by an application of Lemma 4.2.6.

Clearly, $\hat{s}(\underline{\theta}) \geq \frac{1}{2}$ for all $\theta_1 \leq \theta_2 \leq \theta_3$ and $\inf_{\underline{\theta}\in\Omega_o} \hat{s}(\underline{\theta}) = \frac{1}{2}$, which is attained at $\theta_1 = \theta_2 = \theta_3$. We do not know what $\sup_{\underline{\theta}\in\Omega_o} \hat{s}(\underline{\theta}) = \bar{s}$ say, is. However, improvement over $\underline{\delta}_p(\underline{X})$ is possible if $\bar{s} < 1$. The dominating estimator would be $d_{\bar{s}}$. Thus, we have proved the following result.

<u>Theorem 4.2.7</u>: The Pitman estimator $\underline{\delta}_p$ is inadmissible among estimators $\underline{d}_s(\underline{X}) = \underline{X} + s\,\underline{\gamma}(\underline{X})$ if and only if $\bar{s} = \sup_{\underline{\theta}\in\Omega_o} \hat{s}(\underline{\theta}) < 1$. Also estimators $\underline{d}_s(\underline{X})$ for $s < \frac{1}{2}$ are improved by $\underline{d}_{1/2}(\underline{X})$. If $\bar{s} < 1$, the estimator dominating $\underline{\delta}_p$ is $\underline{X} + \bar{s}\,\underline{\gamma}(\underline{X})$.

The technique of Brown (1979) for improving the estimators is described in a subsection of Chapter 2. We employ this technique here to seek an improvement over $\underline{\delta}_p$.

Consider an estimator $\underline{\delta}(\underline{\lambda},\underline{X}) = \underline{\delta}_p(\underline{X}) + \underline{\lambda}(\underline{X})$, where

$$\underline{\lambda}(\underline{x}) = \begin{bmatrix} \lambda_1(\underline{x}) \\ \lambda_2(\underline{x}) \\ \lambda_3(\underline{x}) \end{bmatrix}$$

and let

$$\Delta(\underline{\theta}) = R(\underline{\theta},\underline{\delta}_p) - R(\underline{\theta},\underline{\delta}(\underline{\lambda})).$$

If there is a $\underline{\lambda}$ for which $\Delta(\underline{\theta}) \geq 0$ for all $\theta_1 \leq \theta_2 \leq \theta_3$, the corresponding estimator $\underline{\delta}(\underline{\lambda})$ will be better than $\underline{\delta}_p$. However, it is difficult to guess such a $\underline{\lambda}$ in general. Brown suggested approximating the risk difference by means of Taylor series expansion and obtaining solutions to the approximate equations. In Chapter 2, we developed such an approximate formula for the three-dimensional case. The expression (2.6) on p. 42 obtained there, is of the form

$$\Delta(\underline{\theta}) = -[2(\lambda_{11}(\underline{\theta}) + \lambda_{22}(\underline{\theta}) + \lambda_{33}(\underline{\theta}) - \frac{Q_1(\underline{\theta})}{D(\underline{\theta})}\lambda_1(\underline{\theta}) + \frac{Q_1(\underline{\theta})-Q_2(\underline{\theta})}{D(\underline{\theta})}\lambda_2(\underline{\theta})$$
$$+ \frac{Q_2(\underline{\theta})}{D(\underline{\theta})}\lambda_3(\underline{\theta})) + \lambda_1^2(\underline{\theta}) + \lambda_2^2(\underline{\theta}) + \lambda_3^2(\underline{\theta})] , \qquad (4.2.38)$$

λ_{ij} denoting the first order partial derivative of λ_i with respect to x_j.

If we take $\underline{\lambda}(\underline{x}) = -\frac{a\underline{x}}{|x|^2}$, with $0 < a \leq 2$, which is the choice for James-Stein (1961) estimator, then $\Delta(\underline{\theta}) \geq 0$ for all $\theta_1 \leq \theta_2 \leq \theta_3$. However, the error terms, in reaching the approximation (4.2.38), are complicated and we are not able to show that they are negligible. Still we feel that $\underline{\delta}(\underline{\lambda},\underline{X})$ with $\underline{\lambda}(\underline{x}) = -\frac{a\underline{x}}{|\underline{x}|^2}$ improves upon $\underline{\delta}_p(\underline{X})$.

Another approach to obtain a possible improvement is through empirical Bayes considerations. Efron and Morris (1972a) showed that the James-Stein (1961) estimator of a normal mean vector is, in fact, an empirical Bayes estimator. In subsequent papers, Efron and Morris (1971, 1972b, 1973a) considered various empirical Bayes procedures and showed that these procedures have fairly good risk performances when compared to the usual estimators. Motivated by this fact we obtain an empirical Bayes estimator for the problem of estimating three ordered normal means. We considered a truncated normal prior given by

$$\pi_c(\theta_1,\theta_2,\theta_3) = \frac{1}{\sqrt{c}}\,\varphi\left(\frac{\theta_1}{\sqrt{c}}\right)\frac{1}{\sqrt{c}}\,\varphi\left(\frac{\theta_2}{\sqrt{c}}\right)\frac{1}{\sqrt{c}}\,\varphi\left(\frac{\theta_3}{\sqrt{c}}\right), \quad \theta_1 \le \theta_2 \le \theta_3, \quad c > 0\,. \tag{4.2.39}$$

When c is known, we can easily find the Bayes estimator. However, when it is unknown, we estimate $k^2 = \frac{c}{1+c}$ by $(1 - \frac{a}{|X|^2})^+$, $a > 0$ to get an empirical Bayes estimator

$$\underline{\delta}_a^o(\underline{X}) = \begin{bmatrix} k_o^2 X_1 - \dfrac{Q_{1a}^o(\underline{X})}{D_a^o(\underline{X})} \\[2ex] k_o^2 X_2 + \dfrac{Q_{1a}^o(\underline{X}) - Q_{2a}^o(\underline{X})}{D_a^o(\underline{X})} \\[2ex] k_o^2 X_3 + \dfrac{Q_{2a}^o(\underline{X})}{D_a^o(\underline{X})} \end{bmatrix}, \tag{4.2.40}$$

where $k_o^2 = (1 - \frac{a}{|X|^2})^+$,

$$Q_{1a}^o(x) = \frac{k_o}{\sqrt{2}}\,\varphi\left(\frac{k_o}{\sqrt{2}}(x_1-x_2)\right)\Phi\left(\frac{k_o}{\sqrt{6}}(2x_3-x_1-x_2)\right),$$

$$Q^o_{2a}(\underline{x}) = \frac{k_o}{\sqrt{2}}\,\varphi\left(\frac{k_o}{\sqrt{2}}(x_2-x_3)\right)\,\Phi\left(\frac{k_o}{\sqrt{6}}(x_2+x_3-2x_1)\right),$$

and $$D^o_a(\underline{x}) = P_{\underline{x}}(Z^o_1 \le Z^o_2 \le Z^o_3)$$

with Z^o_i's independently distributed $N(k_o^2 x_i, k_o^2)$ random variables.

For this estimator also the risk expression is too complicated to be compared with the risk of $\underline{\delta}_p$. It would be worth studying this estimator in detail as an alternative to $\underline{\delta}_p$.

4.3 Estimation of Components

Nonminimaxity of the Pitman Estimator for θ_1 and θ_3

Let X_1 and X_2 be independent $N(\theta_1,\tau)$ and $N(\theta_2,1)$ random variables respectively with $\theta_1 \le \theta_2$ and τ a known constant. For estimating θ_2 with squared error loss the component δ_{p2} of the Pitman estimator $\underline{\delta}_p$ has been shown to be minimax (Cohen and Sackrowitz (1970, Theorem 7.1)). This implies the minimaxity of the component for θ_1 as seen below:

The transformations

$$X_1 \to Y_1 = -\frac{X_1}{\sqrt{\tau}}, \quad X_2 \to Y_2 = -\frac{X_2}{\sqrt{\tau}},$$

induce the following transformation on the parameters,

$$\theta_1 \to \eta_1 = -\frac{\theta_1}{\sqrt{\tau}} \quad \text{and} \quad \theta_2 \to \eta_2 = -\frac{\theta_2}{\sqrt{\tau}}.$$

We have $\eta_1 \ge \eta_2$ and $\mathrm{Var}(Y_1) = 1$, $\mathrm{Var}(Y_2) = 1/\tau$. Then the component of the Pitman estimator for estimating η_1 is minimax and is given by

$$\delta^*(Y_1,Y_2) = Y_1 + \left(\frac{1}{\tau}+1\right)^{-1/2} \nu\left(\frac{Y_1-Y_2}{(1+\frac{1}{\tau})^{1/2}}\right)$$

$$= -\frac{1}{\sqrt{\tau}}\left[X_1 - \frac{\tau}{\sqrt{\tau+1}}\,\nu\left(\frac{X_2-X_1}{\sqrt{\tau+1}}\right)\right],$$

where $\nu(.)$ is as defined in Remark 3.3.5.

Now δ^* is minimax for $-\theta_1/\sqrt{\tau}$ if and only if $-\sqrt{\tau}\ \delta^*$ is minimax for θ_1. But $-\sqrt{\tau}\ \delta^*$ is the component δ_{p1} of the Pitman estimator (see (3.3.4)), hence the result.

Notice that this also proves the minimaxity of the Pitman estimator for estimating (θ_1,θ_2) with squared error loss.

Next, let X_1, X_2 and X_3 be independent $N(\theta_1,1)$, $N(\theta_2,1)$ and $N(\theta_3,1)$ random variables with $\theta_1 \leq \theta_2 \leq \theta_3$. We prove that the first and the third components of the Pitman estimator δ_p are not minimax for estimation of θ_1 and θ_3 respectively. This is a departure from what Cohen and Sackrowitz (1970) have proved for $k = 2$ and what Brown (1979, p. 988) asserts about the minimaxity of the last component for any k.

First, we give a lemma to be used in the proof of the non-minimaxity result.

<u>Lemma 4.3.1</u>: Let $D(\underline{x})$, $Q_1(\underline{x})$, $Q_2(\underline{x})$ and $P(\underline{x})$ be as defined in (4.2.23), (4.2.25) and (4.2.29) respectively. Then for all $\underline{x}$, we have

(i) $\frac{1}{6}(x_2 + x_3 - 2x_1)\ Q_2(\underline{x}) + P(\underline{x}) \geq 0$,

(ii) $\frac{1}{6}(2x_3 - x_1 - x_2)\ Q_1(\underline{x}) + P(\underline{x}) \geq 0$,

(iii) $x_2 - x_1 \geq \dfrac{Q_2(\underline{x})-2Q_1(\underline{x})}{D(\underline{x})}$ and

(iv) $x_3 - x_2 \geq \dfrac{Q_1(\underline{x})-2Q_2(\underline{x})}{D(\underline{x})}$.

<u>Proof</u>: (i) $\frac{1}{6}(x_2 + x_3 - 2x_1)\ Q_2(\underline{x}) + P(\underline{x})$

$$= \frac{1}{\sqrt{2}}\,\varphi\left(\frac{x_2-x_3}{\sqrt{2}}\right)\left[\,\frac{1}{6}(x_2+x_3-2x_1)\;\Phi\left(\frac{x_2+x_3-2x_1}{\sqrt{6}}\right) + \frac{1}{\sqrt{6}}\,\varphi\left(\frac{x_2+x_3-2x_1}{\sqrt{6}}\right)\right],$$

which is always nonnegative as $\varphi(y) + y\Phi(y) \geq 0$ for all y. The proof of (ii) is similar. To prove (iii) and (iv), notice that $\underline{\delta}_p$ is the generalized Bayes estimator of $\underline{\theta}$ with respect to the uniform prior restricted to $\theta_1 \leq \theta_2 \leq \theta_3$. Therefore, we have $\delta_{p1} \leq \delta_{p2} \leq \delta_{p3}$, which proves (iii) and (iv).

<u>Theorem 4.3.2</u>: The component δ_{p3} of $\underline{\delta}_p$ is not minimax for estimating θ_3 with respect to squared error loss function.

<u>Proof</u>: From (4.2.26) and (4.2.27),

$$\delta_{p3} = X_3 + \frac{Q_2(\underline{X})}{D(\underline{X})}\,.$$

Define $V_1 = X_1-X_2$, $V_2 = X_3-X_2$ and $V_3 = X_1+X_2+X_3$. Then (V_1, V_2) and V_3 are independently distributed. We have $X_3 = \frac{1}{3}(-V_1+2V_2+V_3)$ and $(X_2+X_3-2X_1) = (V_2-2V_1)$. Also define $\nu_1 = \theta_1-\theta_2$, $\nu_2 = \theta_3-\theta_2$ and $\nu_3 = \theta_1+\theta_2+\theta_3$. Then the risk difference

$$\begin{aligned}\Delta(\underline{\theta}) &= R(\underline{\theta}, X_3) - R(\underline{\theta}, \delta_{p3}) \\ &= E(X_3-\theta_3)^2 - E\left(X_3 + \frac{Q_2(\underline{X})}{D(\underline{X})} - \theta_3\right)^2 \\ &= -\,E\,\frac{Q_2^2(\underline{X})}{D^2(\underline{X})} - 2E(X_3-\theta_3)\,\frac{Q_2(\underline{X})}{D(\underline{X})}\,.\end{aligned}$$

Independence of (V_1, V_2) and V_3 then yields

$$\Delta(\underline{\theta}) = -\,E\,\frac{Q_2^2(\underline{X})}{D^2(\underline{X})} - \frac{2}{3}\,E\,\left\{(2V_2-V_1-2\nu_2+\nu_1)\,\frac{Q_2(\underline{X})}{D(\underline{X})}\right\}$$

Substituting for $E\,\dfrac{Q_2^2(\underline{X})}{D^2(\underline{X})}$ from (4.2.34),

$$\Delta(\underline{\theta}) = E\left[\frac{1}{D(\underline{X})}\left(V_2Q_2(\underline{X}) - P(\underline{X}) + \frac{1}{3}(2V_2-V_1-2\nu_2+\nu_1)\ Q_2(\underline{X})\right)\right]$$

$$- \frac{2}{3}\ E\left[(2V_2-V_1-2\nu_2+\nu_1)\ \frac{Q_2(\underline{X})}{D(\underline{X})}\right]. \qquad (4.3.1)$$

Simplifying (4.3.1) we get

$$\Delta(\underline{\theta}) = -\frac{1}{6}\ E\{(V_2-2V_1)\frac{Q_2(\underline{X})}{D(\underline{X})}\} - E(\frac{P(\underline{X})}{D(\underline{X})}) + \frac{1}{3}(2\theta_3-\theta_1-\theta_2)\ E(\frac{Q_2(\underline{X})}{D(\underline{X})}) .$$

At $\theta_1 = \theta_2 = \theta_3$, the last term in the above expression vanishes whereas the sum of the first two terms is always negative from Lemma 4.3.1(i). Therefore, the risk of δ_{p3} exceeds that of X_3 on $\theta_1 = \theta_2 = \theta_3$. It has been shown in Chapter 3 that X_3 is minimax for estimating θ_3 for $\theta_1 \leq \theta_2 \leq \theta_3$ (Theorem 3.2.3). Hence δ_{p3} is not minimax.

Remark 4.3.1: Changing X_i to $-X_i$ and θ_i to $-\theta_i$, $i = 1,2,3$, it is clear that the component δ_{p1} is not minimax for θ_1. However, we have not been able to resolve the minimaxity question of δ_{p2} for θ_2. Since $\underline{\delta}_p$ is a minimax estimator of $\underline{\theta}$, the proof of Theorem 4.3.2 tells that δ_{p2} has a risk smaller than one in a "neighbourhood" of $\theta_1 = \theta_2 = \theta_3$. One must, however, prove the risk of δ_{p2} to be no larger than one in Ω_o for it to be minimax.

A General Inadmissibility Result

Let X_1 and X_2 be independent normal random variables with means θ_1 and θ_2, $\theta_1 \leq \theta_2$ and variances τ and 1 respectively, τ a known constant. Cohen and Sackrowitz (1970) considered estimators of θ_2 with respect to the squared error loss. They proved that any estimator $\delta(X_2)$, which is inadmissible when $\tau = 1$, remains so when $\tau < 1$. Further, they proved that if $\delta(X_2)$ is such that $\delta(X_2)-X_2$ is bounded above, than $\delta(X_2)$ is inadmissible. By a result of Brown (1971) it implies that any estimator $\delta(X_2)$ with bounded risk is

inadmissible. Cohen and Sackrowitz obtained a class of admissible estimators based on X_2 alone. This result was generalized to a general k. They also considered densities $f(x-\theta_1)$ and $f(x-\theta_2)$, $\theta_1 \leq \theta_2$. When f is symmetric about 0 and loss function is strictly convex, say, $W(|a-\theta_2|)$ with $W(0) = 0$, they proved:

<u>Theorem 4.3.3</u>: If f is such that $P_{\underline{\theta}}(X_1 > X_2) > 0$ for some $\underline{\theta}$ such that $\theta_1 \leq \theta_2$, and if there exists some estimator of θ_2 with finite risk, then X_2 is inadmissible.

The proof of Cohen and Sackrowitz contains a minor error. The expression for A_2+A_3 on p. 2032 should be

$$A_2+A_3 = \int_{-\infty}^{-\eta_2} \int_{-\infty}^{-\eta_2} [W(|t_1+t_2|)-W(|t_1-\eta_2|)+W(|t_2-t_1|)-W(|t_1+\eta_2|)] f(t_1+t_2)\, f(t_1-t_2)dt_2\, dt_1 . \tag{4.3.2}$$

The result can still be proved by modifying arguments of Cohen and Sackrowitz.

When $t_1 < t_2$, the integrand in (4.3.2) can be written as

$$[W(-t_1-t_2)-W(-t_1+\eta_2)+W(t_2-t_1)-W(-t_1-\eta_2)]\ f(t_1+t_2)\ f(t_1-t_2)\ , \tag{4.3.3}$$

since $\eta_2 \geq 0$.

Define $d_1 = t_2-t_1$, $d_2 = -t_1-\eta_2$, $d_3 = -t_1+\eta_2$ and $d_4 = -t_1-t_2$ so that $0 \leq d_1 \leq d_2 \leq d_3 \leq d_4$ and $d_1+d_4 = d_2+d_3$. W being a convex function,

$$W(d_1) + W(d_4) \geq W(d_2) + W(d_3) .$$

Hence the integrand (4.3.3) in A_2+A_3 is nonnegative when $t_1 < t_2$.

When $t_1 > t_2$ and $\eta_2 \geq 0$, the integrand in (4.3.2) becomes

$$W(-t_1-t_2) - W(-t_1+\eta_2) + W(t_1-t_2) - W(-t_1-\eta_2)$$

$$\geq W(-t_1-t_2) - W(-t_2+\eta_2) + W(0) - W(-t_1-\eta_2) \; . \qquad (4.3.4)$$

Take $d_1' = 0$, $d_2' = -t_1-\eta_2$, $d_3' = -t_2+\eta_2$ and $d_4' = -t_1-t_2$ so that

$$0 \leq d_1' \leq d_2' \leq d_3' \leq d_4' \quad \text{and} \quad d_1' + d_4' = d_2' + d_3' \; .$$

Once again convexity of W proves that the right hand side of (4.3.4) is nonnegative. Thus $A_2+A_3 \geq 0$, which proves the result.

The following theorem extends Theorem 4.3.3 and also Theorem 2.1 of Cohen and Sackrowitz.

<u>Theorem 4.3.4</u>: Under the set up of Theorem 4.3.3 the estimator $\delta(X_2) = X_2 + \alpha(X_2)$ is inadmissible for θ_2, if $\alpha(X_2)$ is bounded above by a constant K for which $P_{\underline{\theta}}(X_1 > X_2+2K) > 0$ for some $\underline{\theta}$, $\theta_1 \leq \theta_2$.

<u>Proof</u>: Let $Z_1 = \dfrac{X_1+X_2}{2}$, $Z_2 = \dfrac{X_2-X_1}{2}$,

$$\eta_1 = \frac{\theta_1+\theta_2}{2}, \quad \eta_2 = \frac{\theta_2-\theta_1}{2} \geq 0 \; .$$

In terms of new variables $\delta(X_2)$ is

$$\delta(Z_1, Z_2) = Z_1 + Z_2 + \alpha(Z_1 + Z_2) \; .$$

By hypothesis there is some K such that $\alpha(Z_1+Z_2) \leq K$.

Now consider the estimator

$$\delta^*(Z_1, Z_2) = Z_1 + \max(-K, Z_2) + \alpha(Z_1+Z_2) \; .$$

Then δ^* is a better estimator than δ. The proof is similar to that of Theorem 5.1 of Cohen and Sackrowitz (1970).

Let us consider the risk difference :

$$\Delta = R(\delta,\eta_1,\eta_2) - R(\delta^*,\eta_1,\eta_2)$$

$$= E\ W(|\delta-\theta_2|) - W(|\delta^*-\theta_2|)$$

$$= E[W(|Z_1+Z_2+\alpha(Z_1+Z_2)-\eta_1-\eta_2|) - W(|Z_1+\max(-K,Z_2)+\alpha(Z_1+Z_2)-\eta_1-\eta_2|)]$$

$$= 2\int_{-\infty}^{\infty}\int_{-\infty}^{-K} [W(|z_1+z_2+\alpha(z_1+z_2)-\eta_1-\eta_2|)$$

$$-W(|z_1+\max(-K,z_2)+\alpha(z_1+z_2)-\eta_1-\eta_2|)]$$

$$f(z_1+z_2-\eta_1-\eta_2)\ f(z_1-z_2-\eta_1+\eta_2)\ dz_2\, dz_1 .$$

Making a change of variables $t_1 = z_1-\eta_1$ and $t_2 = z_2-\eta_2$, we get

$$\Delta = 2\int_{-\infty}^{\infty}\int_{-\infty}^{-K-\eta_2} [W(|t_1+t_2+\alpha|)-W(|t_1-K+\alpha-\eta_2|)]\ f(t_1+t_2)\ f(t_1-t_2)dt_2\, dt_1 . \tag{4.3.5}$$

Define subsets S_i of R^2,

$$S_1 = \{(t_1,t_2): |t_1| \le \eta_2,\ t_2 \le -K-\eta_2\},$$

$$S_2 = \{(t_1,t_2): t_1 < -\eta_2,\ t_2 \le -K-\eta_2\},$$

and $S_3 = \{(t_1,t_2): t_1 > \eta_2,\ t_2 \le -K-\eta_2\}$

and let

$$A_i = \int_{S_i} [W(|t_1+t_2+\alpha|)-W(|t_1-K+\alpha-\eta_2|)]\ f(t_1+t_2)\ f(t_1-t_2)dt_2\, dt_1 ,$$

for $i = 1,2,3$.

Then from (4.3.5) $\Delta = 2(A_1 + A_2 + A_3)$.

Now on S_1 we have $-\eta_2 \le t_1 \le \eta_2$, $t_2 \le -\eta_2 - K$, hence $t_1+t_2+\alpha \le 0$ and $t_1-K+\alpha-\eta_2 \le 0$. So

$$W(|t_1+t_2+\alpha|) - W(|t_1-K+\alpha-\eta_2|)$$

$$= W(-t_1-t_2-\alpha) - W(-t_1+K-\alpha+\eta_2) .$$

Now $0 \leq t_1+K-\alpha+\eta_2 \leq -t_1-t_2-\alpha$ and since W is an increasing function, $A_1 \geq 0$.

Making a change of variables $t_1 \to -t_1$ in A_3 and using symmetry of f, we have

$$A_2+A_3 = \int_{-\infty}^{-\eta_2} \int_{-\infty}^{-\eta_2-K} b(t_1,t_2,\alpha,K,\eta_2)\ f(t_1+t_2)\ f(t_1-t_2)dt_2\ dt_1\ , \tag{4.3.6}$$

where

$$b(t_1,t_2,\alpha,K,\eta_2) = W(|t_1+t_2+\alpha|) - W(|t_1-K+\alpha-\eta_2|) + W(|t_2+\alpha-t_1|) - W(|t_1+K-\alpha+\eta_2|)\ . \tag{4.3.7}$$

On S_2, $t_1 < -\eta_2$, $t_2 \leq -\eta_2-K$ and so $t_1+t_2+\alpha \leq 0$ and $t_1+\alpha-K-\eta_2 \leq 0$.

We consider first $t_1 < t_2+\alpha$ and then $t_1 \geq t_2+\alpha$.

(i) If $t_1 < t_2+\alpha$ then $t_1+K-\alpha+\eta_2 \leq t_2+K+\eta_2 \leq 0$. In this case b of (4.3.7) is

$$b = W(-t_1-t_2-\alpha) - W(-t_1+K-\alpha+\eta_2) + W(t_2+\alpha-t_1) - W(-t_1-K+\alpha-\eta_2)$$

$$= W(d_4) - W(d_3) + W(d_1) - W(d_2)\ ,$$

where $d_1 = t_2+\alpha-t_1$, $d_2 = -t_1-K+\alpha-\eta_2$, $d_3 = -t_1+K-\alpha+\eta_2$ and $d_4 = -t_1-t_2-\alpha$. These d_i's satisfy $0 \leq d_1 \leq d_2 \leq d_3 \leq d_4$ and $d_1+d_4 = d_2+d_3 = -2t_1$. Hence by convexity of W we have

$$W(d_4) + W(d_1) \geq W(d_2) + W(d_3)\ .$$

(ii) If $t_1 \geq t_2+\alpha$ then either

$$t_1 \geq \alpha-\eta_2-K \geq t_2+\alpha \tag{4.3.8}$$

or $$\alpha-\eta_2-K > t_1 \geq t_2+\alpha\ . \tag{4.3.9}$$

If (4.3.8) holds then

$$b = W(-t_1-t_2-\alpha) - W(-t_1+K-\alpha+\eta_2) + W(t_1-t_2-\alpha) - W(t_1-\alpha+\eta_2+K) .$$

Now

$$0 \leq -t_1 + K - \alpha + \eta_2 \leq -t_1 - t_2 - \alpha$$

and $0 \leq t_1 - \alpha + \eta_2 + K \leq t_1 - t_2 - \alpha$.

Therefore $b \geq 0$, since W is an increasing function.

If (4.3.9) holds then

$$b = W(-t_1-t_2-\alpha) - W(-t_1+K-\alpha+\eta_2) + W(t_1-t_2-\alpha) - W(-t_1+\alpha-\eta_2-K)$$

$$\geq W(-2t_1) - W(-t_1-\alpha+K+\eta_2) + W(0) - W(-t_1-K+\alpha-\eta_2)$$

$$= W(d_4') - W(d_3') + W(d_1') - W(d_2') ,$$

where

$d_1' = 0$, $d_2' = -t_1-K+\alpha-\eta_2$, $d_3' = -t_1+K-\alpha+\eta_2$ and $d_4' = -2t_1$.

Then $0 = d_1' \leq d_2' \leq d_3' \leq d_4'$ and $d_1' + d_4' = d_2' + d_3' = -2t_1$ and by convexity of W, we have $b \geq 0$.

This completes the proof of the theorem.

CHAPTER - 5

Inadmissibility of 'Affine' Equivariant Estimators

5.1 Introduction

Let X be a random vector taking values in a space $\mathcal{X} \subset R^k$ and have an unknown probability distribution in the family $\mathcal{P} = \{P_\theta: \theta \in \Omega\}$. Assume that $\mathcal{P}$ is dominated by a σ-finite measure μ. The problem is to estimate $h(\theta)$, a measurable function on Ω into R^m. The loss function is taken to be squared error. Let the above estimation problem be invariant under a finite group G of transformations which preserve μ. Under the assumptions that $\tilde{G}$, the group induced by G on the action space $\mathcal{A}$ is commutative and the transformations $\tilde{g}$ in $\tilde{G}$ are linear, Moors (1981, 1985) showed that for each observed x, $\mathcal{A}$ can be curtailed to a subset A_x such that any estimator δ, equivariant under G and satisfying $P_\theta(\delta(X) \in A_X) > 0$ for some θ in Ω, becomes inadmissible. The result leads to substantial reduction of the action space in truncated estimation problems where equivariant estimators taking values on or near the boundary of the parameter space become inadmissible. Moors also gave an estimator which dominates the inadmissible one. In a recent paper, Moors and Van Houwelingen (1987) have proved the above result even when the elements of G do not preserve μ and $\tilde{G}$ is not commutative.

In Section 5.2, we generalize the result of Moors in two directions. First, in place of a finite group, we take G to be a locally compact group. The induced group $\tilde{G}$ is assumed to be commutative and a subgroup of the affine group in place of the

group of linear transformations. The subspaces A_x of $\mathcal{A}$ are defined accordingly. Next, we take the loss function to be any strictly increasing function of the distance between the estimate and the estimand. Sufficient conditions are obtained so that any G-equivariant estimator, taking values outside A_X with positive probability becomes inadmissible. We notice that, for a compact group G, one of the conditions, (5.2.6) always holds. Further when the loss function is squared error, the condition (5.2.11) holds with the compact subset H as G itself. Also, for G compact, we do not require $\tilde{\mathcal{G}}$ to be commutative. However, the condition that the elements of G be measurepreserving cannot be relaxed.

Various applications of the above inadmissibility result are given in Section 5.3. For estimating means θ_1 and θ_2 of two independent normal populations, we take the group of transformations as the group of all the rotations in a plane. Theorem 5.2.7 leads to the conclusion that any G-equivariant admissible estimator and so admissible, must take values in the quadrant, in which the observed x lies. We also consider the estimation of a bounded normal mean and a binomial probability of success with respect to a quartic loss function. The condition (5.2.11) for Theorem 5.2.7 is seen to hold. Estimation of two or more ordered parameters when the underlying distributions are normal or binomial is also taken up.

5.2 The Inadmissibility Result

Let X be a random vector taking values in a space $\mathcal{X} \subset R^k$ and have an unknown probability distribution P_θ, $\theta \in \Omega$, the parameter space Ω a subset of R^p. Let the family of distributions

$\mathcal{P} = \{P_\theta: \theta \in \Omega\}$ be dominated by a σ-finite measure μ and the density of P_θ with respect to μ be $f(x|\theta)$. The problem is to estimate $h(\theta)$, h a measurable function on Ω into R^m. The loss function is

$$L(\theta,a) = W(|h(\theta)-a|) , \tag{5.2.1}$$

where W is a strictly increasing function with $W(0) = 0$.

Without loss of generality, we take the space of possible estimates to be $\mathcal{A}$, the convex closure of $h(\Omega) = \{h(\theta): \theta \in \Omega\}$ (see Lemma 3.3.2 for a justification). We also assume that $h(\Omega)$ contains an open set in R^m.

Let the estimation problem be invariant under a locally compact group G of measurable transformations $g: \mathcal{X} \to \mathcal{X}$, which are measurepreserving, that is,

$$\mu(g^{-1}B) = \mu(B) \tag{5.2.2}$$

for all measurable $B \subset \mathcal{X}$ and all $g \in G$, and let the elements $\tilde{g}$ of $\tilde{G}$, the group induced by G on the action space, be of the form,

$$\tilde{g}(x) = Bx + c , \tag{5.2.3}$$

where B is any nonsingular matrix and c a vector. $\tilde{G}$ is thus, a subgroup of the affine group of transformations. We also assume that $\tilde{G}$ is commutative.

Before giving the inadmissibility result, we state some lemmas. These will be helpful in the proof of the main theorem.

Define a function $\tilde{g}: \mathcal{A} \to \mathcal{A}$ to be isometric, if $\tilde{g}$ preserves the distance between any two points in $\mathcal{A}$, that is,

$$|\tilde{g}(a)-\tilde{g}(b)| = |a-b| \quad \text{for all } a,b \in \mathcal{A} . \tag{5.2.4}$$

Notice that this definition is different from the one considered by Moors (1985), who defines isometry as the property $|\tilde{g}(a)| = |a|$ for all $a \in \mathcal{A}$. However, the statements and the proofs of the lemmas and the main theorem of Moors hold even with this modified definition.

Lemmas 5.2.1-5.2.6 are from Moors and are given for the sake of completeness. The proofs of Lemmas 5.2.1, 5.2.3 and 5.2.4 are modifications of the proofs given by Moors, in view of the fact, that the loss function is any strictly increasing function of the distance and G is a locally compact group of transformations, not necessarily finite.

<u>Lemma 5.2.1</u>: For the G-invariant estimation problem defined above,

(i) $h(\bar{g}(\theta)) = \tilde{g}(h(\theta))$ for all $\theta \in \Omega$, and

(ii) $\tilde{g}$ is isometric,

for each $g \in G$.

<u>Proof</u>: Since the estimation problem is invariant, we should have for $g \in G$,

$$L(\bar{g}(\theta), \tilde{g}(a)) = L(\theta, a)$$

for all $\theta \in \Omega$ and all $a \in \mathcal{A}$. From (5.2.1) this is equivalent to

$$W(|h(\bar{g}(\theta)) - \tilde{g}(a)|) = W(|h(\theta)-a|)$$

for all $\theta \in \Omega$ and all $a \in \mathcal{A}$. Choosing $a = h(\theta)$ in the above relation and using the fact that W is strictly increasing, we get

$$|h(\bar{g}(\theta)) - \tilde{g}(h(\theta))| = 0 \quad \text{for all } \theta \in \Omega,$$

which proves (i). The proof of (ii), now, is the same as in Moors (1985, p. 42).

As a consequence of (i) we have $\bar{G}$ also a commutative group, since $\tilde{G}$ is so.

<u>Lemma 5.2.2</u>: For a G-invariant estimation problem,

$$f(x|\bar{g}(\theta)) = f(g^{-1}(x)|\theta) \quad \text{a.e. with respect to } \mu,$$

for each $g \in G$ and all $\theta \in \Omega$.

<u>Proof</u>: See Moors (1985, p. 43).

Next, we define subsets A_x of $\mathcal{A}$ for each $x \in \mathcal{X}$, as below:

Consider the left invariant Haar measure ν on G (see Chapter 2, p. 37, and also Halmos (1952, p. 254)). This measure is finite on compact subsets of G. Thus we have

$$\nu(H) > 0 \quad \text{for every } H \subset G, \tag{5.2.5}$$

and $\nu(gH) = \nu(H)$ for each $g \in G$ and every $H \subset G$.

We assume that

$$\int_G f(x|\bar{g}(\theta))\, d\nu(g) < \infty . \tag{5.2.6}$$

For each $x \in \mathcal{X}$ and $\theta \in \Omega$, define a probability measure α on G by

$$\alpha(x|\bar{g}(\theta)) = \frac{f(x|\bar{g}(\theta))}{\int_G f(x|\bar{k}(\theta))\, d\nu(k)} \tag{5.2.7}$$

and functions $h_x: \Omega \to \mathcal{A}$ for any fixed $x \in \mathcal{X}$ by

$$\begin{aligned} h_x(\theta) &= \int_G \tilde{g}h(\theta)\, \alpha(x|\bar{g}(\theta))d\nu(g), & \text{if } \int_G f(x|\bar{g}(\theta))d\nu(g) > 0 \\ &= h(\theta) &, \quad \text{if } \int_G f(x|\bar{g}(\theta))d\nu(g) = 0 \end{aligned} \tag{5.2.8}$$

The set A_x is, then defined as the convex closure of $h_x(\Omega) = \{h_x(\theta) : \theta \in \Omega\}$

In Lemmas 5.2.3 and 5.2.4 we state some properties of h_x

Lemma 5.2.3: The functions h_x defined by (5.2.8) satisfy

$$h_{g(x)}(\bar{g}(\theta)) = \tilde{g}(h_x(\theta)) \quad \text{a.e. with respect to } \mu$$

for all $\theta \in \Omega$ and each $g \in G$.

Proof: First, we prove that

$$\alpha(g(x)|\bar{g}(\theta)) = \alpha(x|\theta) \quad \text{a.e. with respect to } \mu \qquad (5.2.9)$$

for all $\theta \in \Omega$ and each $g \in G$.

By definition,

$$\alpha(g(x)|\bar{g}(\theta)) = \frac{f(g(x)|\bar{g}(\theta))}{\int_G f(g(x)|\bar{k}(\theta))\, d\nu(k)}$$

$$= \frac{f(x|\theta)}{\int_G f(x|\bar{g}^{-1}\bar{k}(\theta))d\nu(k)} \quad \text{a.e. with respect to } \mu$$

from Lemma 5.2.2. Since ν is a left invariant measure, transforming $k \to gk$ in the denominator, we get

$$\alpha(g(x)|\bar{g}(\theta)) = \frac{f(x|\theta)}{\int_G f(x|\bar{k}(\theta))\, d\nu(k)} \quad \text{a.e. with respect to } \mu ,$$

and so we have (5.2.9).

Now,

$$h_{g(x)}(\bar{g}(\theta)) = \int_G \tilde{k}\, h(\bar{g}(\theta))\; \alpha(g(x)|\bar{k}\,\bar{g}(\theta))\; d\nu(k)$$

$$\text{if } \int_G f(g(x)|\bar{k}(\bar{g}(\theta))d\nu(k) > 0$$

$$= h\,\bar{g}(\theta) ,$$

$$\text{if } \int_G f(g(x)|\bar{k}\;\bar{g}(\theta))d\nu(k) = 0 .$$

By Lemma 5.2.1(i) and commutativity of $\bar{G}$ and $\tilde{G}$, we get

$$h_{g(x)}(\bar{g}(\theta)) = \int_G \tilde{g}\,\tilde{k}\,h(\theta)\ \alpha(g(x)|\bar{g}\ \bar{k}(\theta))d\nu(k)\ ,$$

$$\text{if } \int_G f(gx|\bar{g}\ \bar{k}(\theta))d\nu(k) > 0$$

$$= \tilde{g}\ h(\theta)\ ,$$

$$\text{if } \int_G f(g(x)|\bar{g}\ \bar{k}(\theta))d\nu(k) = 0\ .$$

Since $\tilde{g}$ is an affine transformation, α satisfies (5.2.9) and $\alpha(x|\bar{g}(\theta))d\nu(g)$ is a probability measure on G, we get the result.

<u>Lemma 5.2.4</u>: For sets A_x defined above

(i) $A_x \subset \mathcal{A}$ for all $x \in \mathcal{X}$, and

(ii) $A_{g(x)} = \tilde{g}(A_x)$ for any $g \in G$.

<u>Proof</u>: To prove (i), we notice the result that if S is a convex subset of R^m and Z is an m-dimensional random vector with $P(Z \in S) = 1$ and $E(Z) < \infty$, then $E(Z) \in S$ (see, for example, Ferguson (1967, p. 74)). When $\int_G f(x|\bar{g}(\theta))d\nu(g) > 0$, $h_x(\theta)$ can be written as $E\ \tilde{g}\ h(\theta)$ where the expectation is taken over g with respect to a probability density $\alpha(x|\bar{g}(\theta))d\nu(g)$. Also $P(\tilde{g}\ h(\theta) \in \mathcal{A}) = 1$ and $\mathcal{A}$ is convex so that $h_x(\theta) \in \mathcal{A}$. Since A_x is the smallest convex set containing $h_x(\Omega)$, $A_x \subset \mathcal{A}$.

For a proof of (ii), see Moors (1985, p. 44).

Let S be a closed convex subset of R^m. The projection of a point $x \in R^m$ on S is defined to be the unique point $x_o \in S$ such that

$$|x_o - x|^2 = \operatorname*{Inf}_{x \in S} |s-x|^2\ .$$

Lemmas 5.2.5 and 5.2.6 below state some properties of the projections. See Moors (1985, pp. 45-46) for their proofs.

Lemma 5.2.5: Let S be a closed convex subset of R^m and let x_o be the projection of a point $x \in R^m$ on S. If $x \notin S$,

$$|x-s|^2 > |x_o-s|^2 \quad \text{for all } s \in S .$$

Lemma 5.2.6: Let $S \subset R^m$ be closed and convex and the projection of $x \in R^m$ on S be x_o. If $\tilde{g}$ is isometric, the projection of $\tilde{g}(x)$ on $\tilde{g}(S)$ is given by $\tilde{g}(x_o)$.

Now, we state and prove the main theorem of this section.

Theorem 5.2.7: Consider a G-invariant estimation problem with G a locally compact group, the induced group $\tilde{G}$ a commutative subgroup of the affine group of transformations and the loss function (5.2.1). Assume that the conditions (5.2.2) and (5.2.6) are satisfied. Then any G-equivariant estimator d, for which

$$P_{\theta'}(d(X) \notin A_X) > 0 \quad \text{for some } \theta' \in \Omega , \tag{5.2.10}$$

is dominated by d_o, where $d_o(x)$ is the projection of $d(x)$ on A_x, provided there exists a compact subset H of G such that

$$\int_H [W(|\tilde{g}\, h(\theta)-d(x)|) - W(|\tilde{g}\, h(\theta)-d_o(x)|)]\, f(x|\bar{g}(\theta'))\, d\nu(g) > 0 , \tag{5.2.11}$$

for all $x \in B_d = \{x: d(x) \notin A_x\}$.

Proof: The measure ν is finite on compact subsets of G and so without loss of generality we can write $\nu(H) = 1$. Consider the risk function of d:

$$R(\theta',d) = \int_{\mathfrak{X}} W(|h(\theta')-d(x)|)\, f(x|\theta')\, d\mu(x) .$$

Since d is equivariant, the risk of d is constant on the orbits of θ'. This implies

$$R(\theta',d) = \int_H R(\bar{g}(\theta'),d)\, d\nu(g)$$

$$= \int_H \int_{\mathfrak{X}} W(|h\,\bar{g}(\theta')-d(x)|)\, f(x|\bar{g}(\theta'))\, d\mu(x)\, d\nu(g) .$$

Lemma 5.2.1(i) and an interchange in the order of integration gives

$$R(\theta',d) = \int_{\mathfrak{X}} \int_H W(|\tilde{g}\,h(\theta')-d(x)|)\, f(x|\bar{g}(\theta'))\, d\nu(g)\, d\mu(x) . \tag{5.2.12}$$

It can be shown easily that d_o is also a G-equivariant estimator (see Moors (1985, p. 46)): By Lemma 5.2.6, the projection of $\tilde{g}(d(x))$ on $\tilde{g}(A_x)$ is $\tilde{g}(d_o(x))$. By equivariance of d and Lemma 5.2.4 we have $\tilde{g}(d(x)) = d(g(x))$ and $\tilde{g}(A_x) = A_{g(x)}$ and so $\tilde{g}(d_o(x)) = d_o(g(x))$. Hence an expression of the form (5.2.12) is valid for $R(\theta',d_o)$ also. Thus,

$$R(\theta',d)-R(\theta',d_o)$$

$$= \int_{\mathfrak{X}} \int_H [W(|\tilde{g}\,h(\theta')-d(x)|) - W(|\tilde{g}\,h(\theta')-d_o(x)|)]\, f(x|\bar{g}(\theta'))d\nu(g)d\mu(x)$$

$$= \int_{B_d} \int_H [W(|\tilde{g}\,h(\theta')-d(x)|)- W(|\tilde{g}\,h(\theta')-d_o(x)|)]\, f(x|\bar{g}(\theta'))d\nu(g)d\mu(x) .$$

which is positive by (5.2.10) and (5.2.11). This completes the proof of the theorem.

<u>Remark 5.2.1</u>: The sufficient condition (5.2.11) above will be satisfied if for each $x \in B_d$ there exists a $y_{x,\theta'} \in A_x$ such that

$$\int_H [W(|\tilde{g}\,h(\theta')-d(x)|) - W(|\tilde{g}\,h(\theta')-d_o(x)|)]\, \alpha^*(x|\bar{g}(\theta'))d\nu(g)$$
$$\geq W(|y_{x,\theta'}-d(x)|) - W(|y_{x,\theta'}-d_o(x)|) , \tag{5.2.13}$$

where $\alpha^*(x \mid \bar{g}(\theta')) = \dfrac{f(x \mid \bar{g}(\theta'))}{\int_H f(x \mid \bar{g}(\theta'))d\nu(g)}$.

By definition $x \in B_d \Rightarrow d(x) \notin A_x$ and so by Lemma 5.2.5,

$$|y_{x,\theta'} - d(x)| > |y_{x,\theta'} - d_o(x)| .$$

Since W is a strictly increasing function, the right hand side of (5.2.13) is positive for all $x \in B_d$ and so (5.2.11) holds.

Remark 5.2.2: When the group G is compact, the left invariant measure ν on G is also right invariant and finite. The condition (5.2.6) is satisfied. In this case Lemma 5.2.3 can be proved without using commutativity of $\bar{G}$ or $\tilde{G}$ as shown below:

$$h_{g(x)}(\bar{g}(\theta)) = \int_G \tilde{k}\, h(\bar{g}(\theta))\ \alpha(g(x) \mid \bar{k}(\bar{g}(\theta)))\ d\nu(k) ,$$

$$\text{if} \int_G f(g(x) \mid \bar{k}\, \bar{g}(\theta))\ d\nu(k) > 0 ,$$

$$= h\, \bar{g}(\theta) ,$$

$$\text{if} \int_G f(g(x) \mid \bar{k}\, \bar{g}(\theta))\ d\nu(k) = 0 .$$

Using Lemma 5.2.1(i) and identity (5.2.9), we can write

$$h_{g(x)}(\bar{g}(\theta)) = \int_G \tilde{g}\, \tilde{g}^{-1}\, \tilde{k}\, \tilde{g}\, h(\theta)\ \alpha(x \mid \bar{g}^{-1}\, \bar{k}\, \bar{g}(\theta))\ d\nu(k) ,$$

$$\text{if} \int_G f(x \mid \bar{g}^{-1}\, \bar{k}\, \bar{g}(\theta))\ d\nu(k) > 0$$

$$= \tilde{g}\, h(\theta) ,$$

$$\text{if} \int_G f(x \mid \bar{g}^{-1}\, \bar{k}\, \bar{g}(\theta))\ d\nu(k) = 0$$

a.e. with respect to μ.

We transform $k \to g.k\, g^{-1}$ in the above integrals and use the fact that ν is both left and right invariant to get

$$h_{g(x)}(\bar{g}(\theta)) = \int_G \tilde{g}\,\tilde{k}\,h(\theta)\ \alpha(x \mid \bar{k}(\theta))\ d\nu(k)\ , \qquad \text{if } \int_G f(x \mid \bar{k}(\theta))\ d\nu(k) > 0$$

$$= \tilde{g}\,h(\theta)\ , \qquad \text{if } \int_G f(x \mid \bar{k}(\theta))\ d\nu(k) = 0$$

a.e. with respect to μ. Since $\tilde{g}$ is an affine transformation and $\int_G \alpha(x \mid \bar{k}(\theta))d\nu(k) = 1$, the proof of the lemma is complete.

In the condition (5.2.11) the set H can now be taken to be G itself. In the proof of the theorem we can start with

$$R(\theta',d) = \int_G R(\bar{g}(\theta'),d)\ d\nu(g)\ .$$

Remark 5.2.3: When G is compact and the loss function is squared error, we can easily verify that (5.2.11) holds. The left hand side of (5.2.11) with H = G is

$$\int_G (\,|\tilde{g}\,h(\theta')-d(x)|^2 - |\tilde{g}\,h(\theta')-d_o(x)|^2)\ f(x \mid \bar{g}(\theta'))\ d\nu(g)$$

$$= (|d(x)|^2 - |d_o(x)|^2) \int_G f(x \mid \bar{g}(\theta'))d\nu(g)$$
$$\qquad + 2(d_o(x)-d(x))^T \int_G \tilde{g}\ h(\theta')f(x \mid \bar{g}(\theta'))d\nu(g)\ ,$$

where y^T denotes transpose of the vector y. By definition (5.2.8) we can write the above expression as

$$= [|d(x)|^2 - |d_o(x)|^2 + 2(d_o(x)-d(x))^T\ h_x(\theta')](\int_G f(x \mid \bar{g}(\theta'))d\nu(g))$$

$$= (|h_x(\theta')-d(x)|^2 - |h_x(\theta')-d_o(x)|^2)(\int_G f(x \mid \bar{g}(\theta'))d\nu(g))\ . \tag{5.2.14}$$

This last expression (5.2.14) is positive by Lemma 5.2.5, whenever $x \in B_d$ and this proves our claim.

<u>Remark 5.2.4</u>: When G is locally compact and the loss function is squared error, a sufficient condition for (5.2.11) to hold is that there exists a compact set $H \subset G$ for which $h_x^*(\theta') \in A_x$ for all $x \in B_d$, where

$$h_x^*(\theta') = \int_H \tilde{g}\, h(\theta')\, \alpha^*(x \mid \bar{g}(\theta'))d\nu(g) .$$

Proceeding as in the previous remark we get the left hand side of (5.2.11) as

$$|h_x^*(\theta')-d(x)|^2 - |h_x^*(\theta')-d_o(x)|^2 ,$$

which is positive by Lemma 5.2.5 for all $x \in B_d$.

<u>Remark 5.2.5</u>: An alternative to the condition (5.2.11), under which the result of Theorem 5.2.7 holds, is:

There exists a probability measure ν^* on G such that

$$\int_G [W(|\tilde{g}\, h(\theta')-d(x)|) - W(|\tilde{g}\, h(\theta')-d_o(x)|)]\ f(x \mid \bar{g}(\theta'))d\nu^*(g) > 0$$

for all $x \in B_d$.

For a proof start with $R(\theta',d) = \int_G R(\bar{g}(\theta'),d)d\nu^*(g)$ and proceed as in the proof of Theorem 5.2.7.

<u>Remark 5.2.6</u>: The general set up of G a locally compact group (not necessarily affine) such that $\tilde{G}$ is a subgroup of the affine group is probably only of theoretical interest in view of the fact that μ is usually Lebesgue or counting measure. The requirement that g preserves μ severely limits the choice of G. We have not been able to find an example where $g \in G$ is measurepreserving and

not affine but $\tilde{g} \in \tilde{G}$ is affine. Moors (1985) also in his examples takes G itself to be a subgroup of the affine group.

5.3 Applications to Truncated Estimation Problems

Example 5.3.1 (Estimating a restricted binomial parameter): Let X be a binomial random variable with parameters n and θ, $\theta \in (1-P,P)$, $\frac{1}{2} < P < 1$ and let the loss in estimating θ by a be

$$L(\theta,a) = (\theta-a)^2 .$$

The problem is invariant under the group $G = \{e,g\}$, where $g(x) = n-x$ and e is the identity element. G induces groups $\bar{G} = \{\bar{e},\bar{g}\}$ and $\tilde{G} = \{\tilde{e},\tilde{g}\}$ on Ω and $\mathcal{A}$ respectively where $\bar{g}(\theta) = 1-\theta$ and $\tilde{g}(a) = 1-a$. A G-equivariant estimator $d(x)$ satisfies $d(x)+d(n-x) = 1$. This problem was discussed by Moors (1985) in connection with randomized response models for $n = 1$.

We have

$$f(x\,|\,\theta) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \quad x = 0,1,2,\ldots,n$$

and so α and h_x of (5.2.7) and (5.2.8) are

$$\alpha(x\,|\,\theta) = \frac{\theta^x(1-\theta)^{n-x}}{\theta^x(1-\theta)^{n-x} + (1-\theta)^x\theta^{n-x}},$$

$$\alpha(x\,|\,\bar{g}(\theta)) = \frac{(1-\theta)^x\theta^{n-x}}{\theta^x(1-\theta)^{n-x} + (1-\theta)^x\theta^{n-x}}$$

and

$$\begin{aligned} h_x(\theta) &= \theta\,\alpha(x\,|\,\theta) + (1-\theta)\alpha(x\,|\,\bar{g}(\theta)) \\ &= \frac{\theta^{x+1}(1-\theta)^{n-x} + (1-\theta)^{x+1}\theta^{n-x}}{\theta^x(1-\theta)^{n-x} + (1-\theta)^x\theta^{n-x}}, \end{aligned}$$

$$x = 0,1,\ldots,n \text{ and } \theta \in (1-P,P)$$

To determine A_x we consider the following three cases: (i) $x < \frac{n}{2}$, (ii) $x = \frac{n}{2}$ and (iii) $x > \frac{n}{2}$. When $x < \frac{n}{2}$, let $r = \frac{n}{2} - x$ and so

$$h_x(\theta) = \frac{\theta(1-\theta)^{2r} + (1-\theta)\theta^{2r}}{(1-\theta)^{2r} + \theta^{2r}}, \quad \theta \in (1-P,P) .$$

It can be checked easily that $h_x(\theta) \leq \frac{1}{2}$ for all $\theta \in (1-P,P)$ with equality occurring when $\theta = \frac{1}{2}$ and so $\sup_{\theta \in (1-P,P)} h_x(\theta) = \frac{1}{2}$. To determine the minimum value of $h_x(\theta)$ we observe that $\frac{\partial}{\partial\theta} h_x(\theta)$ is positive for $\theta < \frac{1}{2}$ and negative for $\theta > \frac{1}{2}$, that is, $h_x(\theta)$ increases for θ varying from $1-P$ to $\frac{1}{2}$ and decreases thereafter. Therefore $h_x(\theta)$ attains its minimum at $\theta = P$ or $1-P$ giving

$$\inf_{\theta \in (1-P,P)} h_x(\theta) = \frac{(1-P)P^{2r} + P(1-P)^{2r}}{P^{2r} + (1-P)^{2r}} .$$

Hence when $x < \frac{n}{2}$, $A_x = [\frac{(1-P)P^{n-2x} + P(1-P)^{n-2x}}{P^{n-2x} + (1-P)^{n-2x}}, \frac{1}{2}]$.

In case (ii), $x = \frac{n}{2}$ and $h_x(\theta)$ simply reduces to $\frac{1}{2}$ for every θ. Thus $A_x = \{\frac{1}{2}\}$. Case (iii) can be treated as Case (i) and we get

$$A_x = [\frac{1}{2}, \frac{P^{2x-n+1} + (1-P)^{2x-n+1}}{P^{2x-n} + (1-P)^{2x-n}}] \quad \text{for } x > \frac{n}{2} .$$

By Theorem 5.2.7 any admissible equivariant estimator $d(X)$ should satisfy

$$d(x) \in [\frac{(1-P)P^{n-2x} + P(1-P)^{n-2x}}{P^{n-2x} + (1-P)^{n-2x}}, \frac{1}{2}], \text{ if } x < \frac{n}{2}; \ d(x) = \frac{1}{2},$$

if $x = \frac{n}{2}$ and $d(x) \in [\frac{1}{2}, \frac{(1-P)^{2x-n+1} + P^{2x-n+1}}{(1-P)^{2x-n} + P^{2x-n}}]$, if $x > \frac{n}{2}$.

<u>Example 5.3.2 (Estimating a restricted binomial parameter, $n = 1$)</u>: Let X be a Bernoulli random variable with θ the probability of success, $\theta \in (1-P,P)$, $\frac{1}{2} < P < 1$. Let the loss in estimating θ be

$$L(\theta, a) = (\theta-a)^4 . \tag{5.3.1}$$

The problem is invariant under the group $G = \{e,g\}$, where $g(x) = 1-x$ and e is the identity element. G induces groups $\bar{G} = \{\bar{e},\bar{g}\}$ and $\tilde{G} = \{\tilde{e},\tilde{g}\}$ on Ω and $\mathcal{A}$ respectively, where $\bar{g}(\theta) = 1-\theta$ and $\tilde{g}(a) = 1-a$. A G-equivariant estimator $d(x)$ satisfies $d(x)+d(1-x) = 1$. We obtain subsets A_x of $\mathcal{A}$ and show that the condition (5.2.11) of Theorem 5.2.7 holds with $H = G$, ν the counting measure and the loss function given by (5.3.1).

Taking $n = 1$ in Example 5.3.1 the subsets A_x of $\mathcal{A}$ are

$$A_o = [2P(1-P), \tfrac{1}{2}] \tag{5.3.2}$$

and

$$A_1 = [\tfrac{1}{2}, P^2+(1-P)^2]. \tag{5.3.3}$$

We now verify condition (5.2.11) with $H = G$. Let $x = 0$ and $d(X)$ be an equivariant estimator with $d(0) \notin A_o$. Let us first consider the case $d(0) > \frac{1}{2}$. Then the projection $d_o(0)$ of $d(0)$ on A_o is $\frac{1}{2}$. Writing $a = d(0)$ the condition (5.2.11) takes the form

$$((\theta-a)^4 - (\theta-\tfrac{1}{2})^4)(1-\theta) + ((1-\theta-a)^4 - (1-\theta-\tfrac{1}{2})^4)\theta > 0 \tag{5.3.4}$$

for $a > \frac{1}{2}$ and $\theta \in (1-P,P)$. Denote the left hand side of the inequality (5.3.4) by $P_1(a,\theta)$. Then $P_1(\frac{1}{2}, \theta) = 0$. Consider

$$\frac{\partial P_1(a,\theta)}{\partial a} = - 4(\theta-a)^3(1-\theta) - 4(1-\theta-a)^3\theta$$

and

$$\frac{\partial^2 P_1(a,\theta)}{\partial a^2} = 12(\theta-a)^2(1-\theta) + 12(1-\theta-a)^2\theta .$$

Now at $a = \frac{1}{2}$, $\frac{\partial P_1(a,\theta)}{\partial a} = 4(\frac{1}{2}-\theta)^3(1-2\theta) > 0$ for all $\theta \in (1-P,P)$ and $\frac{\partial^2 P_1(a,\theta)}{\partial a^2} > 0$ for all $\theta \in (1-P,P)$ and a. Thus $\frac{\partial P_1}{\partial a} > 0$ for $a > \frac{1}{2}$ and so $P_1(a,\theta) > 0$ for $a > \frac{1}{2}$ and $\theta \in (1-P,P)$, thus verifying the condition (5.3.4).

Next, we take up $d(0) < 2P(1-P)$. In this case $d_o(0) = 2P(1-P)$. Once again writing a for d(0) the condition (5.2.11) reduces to

$$((\theta-a)^4-(\theta-2P(1-P))^4)(1-\theta) + ((1-\theta-a)^4-(1-\theta-2P(1-P))^4)\theta > 0 \tag{5.3.5}$$

for $a < 2P(1-P)$ and $\theta \in (1-P,P)$. Denote the left hand side of the inequality (5.3.5) by $P_2(a,\theta)$. We notice that $P_2(2P(1-P),\theta) = 0$ and

$$\frac{\partial P_2(a,\theta)}{\partial a} = -4(\theta-a)^3(1-\theta) - 4(1-\theta-a)^3\theta$$

$$\frac{\partial^2 P_2(a,\theta)}{\partial a^2} = 12(\theta-a)^2(1-\theta) + 12(1-\theta-a)^2\theta .$$

Since $\frac{\partial^2 P_2(a,\theta)}{\partial a^2} > 0$ for any a and all $\theta \in (1-P,P)$, we have $\frac{\partial P_2(a,\theta)}{\partial a}$ an increasing function of a. Thus we will be through if we can show that at $a = 2P(1-P)$ the value of $\frac{\partial P_2(a,\theta)}{\partial a}$ is negative for all $\theta \in (1-P,P)$. For convenience let us denote this value by $-4Q(\theta)$. Then

$$Q(\theta) = (\theta-2P(1-P))^3(1-\theta) + (1-\theta-2P(1-P))^3\theta .$$

It can be shown easily that $Q(\theta)$ attains its minimum value in $(1-P,P)$ either at $\theta = \frac{1}{2}$ or at $\theta = P, 1-P$. Now $Q(\frac{1}{2}) = (\frac{1}{2} - 2P(1-P))^3 > 0$ for all $P \in (\frac{1}{2}, 1)$ and $Q(P) = Q(1-P) = P(1-P)(2P-1)^4 > 0$. So $Q(\theta) > 0$ for $\theta \in (1-P,P)$. This completes the verification of the condition (5.3.5).

In a similar way, one can verify that (5.2.11) holds when $x = 1$ and $d(1) \notin A_1$. Thus Theorem 5.2.7 is applicable here. Any admissible equivariant estimator d must satisfy $d(0) \in [2P(1-P), \frac{1}{2}]$ and $d(1) \in [\frac{1}{2}, 1-2P+2P^2]$.

<u>Example 5.3.3 (Estimating two ordered binomial probabilities of success)</u>: Let X_1 and X_2 be independent $\mathrm{Bin}(n,\theta_1)$ and $\mathrm{Bin}(n,\theta_2)$ random variables and $\theta_1 \le \theta_2$. Consider estimation of $\underline{\theta} = (\theta_1,\theta_2)$ with the loss function the sum of squared errors:

$$L(\underline{\theta},\underline{a}) = |\underline{\theta}-\underline{a}|^2 . \tag{5.3.6}$$

The estimation problem is invariant under group $G = \{e,g\}$ where $g(x_1,x_2) = (n-x_2, n-x_1)$. G induces groups $\bar{G} = \{\bar{e},\bar{g}\}$ and $\tilde{G} = \{\tilde{e},\tilde{g}\}$ on Ω and $\mathcal{A}$ respectively, where $\bar{g}(\underline{\theta}) = (1-\theta_2, 1-\theta_1)$ and $\tilde{g}(\underline{a}) = (1-a_2, 1-a_1)$. Thus we have $\tilde{g}(\underline{a}) = B\underline{a}+\underline{c}$, where $B = \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}$ and $\underline{c} = (1,1)$.

Here,

$$f(\underline{x} \mid \underline{\theta}) = \binom{n}{x_1} \theta_1^{x_1}(1-\theta_1)^{n-x_1} \binom{n}{x_2} \theta_2^{x_2}(1-\theta_2)^{n-x_2} ,$$

$$x_i = 0,1,\ldots,n, \; i = 1,2; \; 0 \le \theta_1 \le \theta_2 \le 1 .$$

We get

$$\alpha(\underline{x} \mid \underline{\theta}) = \frac{\theta_1^{x_1}(1-\theta_1)^{n-x_1}\theta_2^{x_2}(1-\theta_2)^{n-x_2}}{\theta_1^{x_1}(1-\theta_1)^{n-x_1}\theta_2^{x_2}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1}\theta_2^{n-x_1}(1-\theta_1)^{x_2}\theta_1^{n-x_2}}$$

and

$$\alpha(\underline{x} \mid \bar{g}(\underline{\theta})) = \frac{(1-\theta_2)^{x_1}\theta_2^{n-x_1}(1-\theta_1)^{x_2}\theta_1^{n-x_2}}{\theta_1^{x_1}(1-\theta_1)^{n-x_1}\theta_2^{x_2}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1}\theta_2^{n-x_1}(1-\theta_1)^{x_2}\theta_1^{n-x_2}}$$

Therefore,

$$h_{\underline{x}}(\underline{\theta}) = \begin{bmatrix} \dfrac{\theta_1^{x_1+1}(1-\theta_1)^{n-x_1}\theta_2^{x_2}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1+1}(1-\theta_1)^{x_2}\theta_2^{n-x_1}\theta_1^{n-x_2}}{\theta_1^{x_1}\theta_2^{x_2}(1-\theta_1)^{n-x_1}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1}(1-\theta_1)^{x_2}\theta_2^{n-x_1}\theta_1^{n-x_2}} \\[2ex] \dfrac{\theta_1^{x_1}\theta_2^{x_2+1}(1-\theta_1)^{n-x_1}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1}(1-\theta_1)^{x_2+1}\theta_2^{n-x_1}\theta_1^{n-x_2}}{\theta_1^{x_1}\theta_2^{x_2}(1-\theta_1)^{n-x_1}(1-\theta_2)^{n-x_2} + (1-\theta_2)^{x_1}(1-\theta_1)^{x_2}\theta_2^{n-x_1}\theta_1^{n-x_2}} \end{bmatrix}$$

Next we determine $A_{\underline{x}}$. Clearly, both the components lie between 0 and 1. As $\theta_1 \leq \theta_2$, we have the first component in $h_{\underline{x}}(\underline{\theta})$ always less than or equal to the second component. There are three cases under consideration:

(i) $x_1 + x_2 = n$,

(ii) $x_1 + x_2 < n$, and

(iii) $x_1 + x_2 > n$.

In case (i), using $x_2 = n-x_1$ we get

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{2}\begin{bmatrix} 1+\theta_1-\theta_2 \\ 1+\theta_2-\theta_1 \end{bmatrix}$$

and so $A_{\underline{x}}$ is the line segment joining the points $(0,1)$ and $(\frac{1}{2}, \frac{1}{2})$ in the two dimensional co-ordinate plane. In case (ii), we write $m = n-x_1-x_2$. Then

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{(1-\theta_1)^m(1-\theta_2)^m+\theta_1^m\theta_2^m}\begin{bmatrix} \theta_1(1-\theta_2)((1-\theta_1)^m(1-\theta_2)^{m-1}+\theta_1^{m-1}\theta_2^m) \\ \theta_2(1-\theta_1)((1-\theta_1)^{m-1}(1-\theta_2)^m+\theta_1^m\theta_2^{m-1}) \end{bmatrix}.$$

The sum of the two components of $h_{\underline{x}}(\underline{\theta})$ is

$$w(\theta_1,\theta_2) = \frac{(1-\theta_1)^m(1-\theta_2)^m(\theta_1+\theta_2) + \theta_1^m\theta_2^m(2-\theta_1-\theta_2)}{(1-\theta_1)^m(1-\theta_2)^m + \theta_1^m\theta_2^m} .$$

Now $w(\theta_1,\theta_2) \le 1$ is equivalent to

$$((1-\theta_1)^m(1-\theta_2)^m-\theta_1^m\theta_2^m)(1-\theta_1-\theta_2) \ge 0 ,$$

which is true as $0 \le \theta_1, \theta_2 \le 1$. Also points $\underline{\theta} = (0,0)$, $(\frac{1}{2}, \frac{1}{2})$ and $(0,1)$ correspond to the same points $h_{\underline{x}}(\underline{\theta})$ in the plane and so A_x is the triangle with sides given by the lines $x = 0$, $x+y = 1$ and $x = y$. In a similar way for case (iii), A_x is the triangle with the sides given by the lines $y = 1$, $x+y = 1$ and $x = y$. Thus we see that there is a substantial reduction in the action space. Theorem 5.2.7 is applicable as we have a finite group and the loss function is squared error. Any admissible equivariant estimator must lie in A_x with probability one.

Example 5.3.4 (Estimating bounded normal mean): Let X be a normal random variable with mean θ and variance unity, $\theta \in [-m,m]$. Also let the loss in estimating θ be (5.3.1). This estimation problem is invariant under $G = \{e,g\}$, where $g(x) = -x$. The induced groups are $\bar{G} = \{\bar{e},\bar{g}\}$ and $\tilde{G} = \{\tilde{e},\tilde{g}\}$, where $\bar{g}(\theta) = -\theta$ and $\tilde{g}(a) = -a$. A G-equivariant estimator satisfies $d(-x) = -d(x)$. Then h_x as obtained in Moors (1985, p. 49) is

$$h_x(\theta) = \frac{\theta\, \varphi(x-\theta) - \theta\, \varphi(x+\theta)}{\varphi(x-\theta) + \varphi(x+\theta)} = \theta \tanh(\theta x) .$$

Thus we have

$$\begin{aligned} A_x &= [0,\ m \tanh(mx)\,], && \text{if } x > 0 \\ &= [m \tanh(mx),\ 0\,], && \text{if } x < 0 \qquad (5.3.7) \\ &= \{0\} , && \text{if } x = 0 . \end{aligned}$$

Now we verify condition (5.2.11) of Theorem 5.2.7. Let us consider the case $x > 0$ and $d(x) < 0$. Then $d_o(x) = 0$ and writing a for $d(x)$ and G for H the condition reduces to

$$((\theta-a)^4 - \theta^4)e^{\theta x} + ((\theta+a)^4 - \theta^4)e^{-\theta x} > 0, \tag{5.3.8}$$

Denote the left hand side of the inequality (5.3.8) by $P_3(a,\theta)$. Then $P_3(0,\theta) = 0$ and

$$\frac{\partial P_3}{\partial a} = -4(\theta-a)^3 e^{\theta x} + 4(\theta+a)^3 e^{-\theta x},$$

$$\frac{\partial^2 P_3}{\partial a^2} = 12(\theta-a)^2 e^{\theta x} + 12(\theta+a)^2 e^{-\theta x}.$$

At $a = 0$, $\frac{\partial P_3}{\partial a} = -4\theta^3(e^{\theta x} - e^{-\theta x})$ which is negative for all θ and all $x > 0$. Also $\frac{\partial^2 P_3}{\partial a^2} > 0$ for all θ and all a, so that $\frac{\partial P_3}{\partial a} < 0$ for $a < 0$. This proves that $P_3(a,\theta) > 0$ for all θ and $a < 0$. Hence the inequality (5.3.8) is satisfied for all $a < 0$ and $\theta \in [-m,m]$. In a similar manner, we can verify (5.2.11) for the other cases. Thus Theorem 5.2.7 is applicable and any admissible G-equivariant estimator must have $d(X) \in A_X$ with probability one.

<u>Example 5.3.5 (Estimating k ordered normal means)</u>: Let $X_1, X_2, \ldots, X_k$ be independent normal random variables with means $\theta_1, \theta_2, \ldots, \theta_k$, $\theta_1 \leq \theta_2 \leq \cdots \leq \theta_k$ and common variance unity. We estimate $\underline{\theta} = (\theta_1, \theta_2, \ldots, \theta_k)$ with the squared error loss function:

$$L(\underline{\theta}, \underline{a}) = |\underline{\theta} - \underline{a}|^2 .$$

This estimation problem is invariant under the group $G = \{e, g\}$, where $g((x_1, x_2, \ldots, x_k)) = (-x_k, -x_{k-1}, \ldots, -x_1)$. The group G induces groups $\bar{G} = \{\bar{e}, \bar{g}\}$ and $\tilde{G} = \{\tilde{e}, \tilde{g}\}$ on Ω and $\mathcal{A}$ respectively with $\bar{G}$ and $\tilde{G}$ the same as G.

Now

$$f(\underline{x} \mid \underline{\theta}) = \varphi(x_1-\theta_1) \ldots \varphi(x_k-\theta_k) ,$$

$$\alpha(\underline{x} \mid \underline{\theta}) = \frac{\varphi(x_1-\theta_1) \ldots \varphi(x_k-\theta_k)}{\varphi(x_1-\theta_1) \ldots \varphi(x_k-\theta_k) + \varphi(x_1+\theta_k) \ldots \varphi(x_k+\theta_1)} = \frac{1}{1+e^{-y}} ,$$

and

$$\alpha(\underline{x} \mid \bar{g}(\underline{\theta})) = \frac{\varphi(x_1+\theta_k)\ldots\varphi(x_k+\theta_1)}{\varphi(x_1-\theta_1)\ldots\varphi(x_k-\theta_k) + \varphi(x_1+\theta_k)\ldots\varphi(x_k+\theta_1)} = \frac{1}{1+e^{y}} ,$$

where $y = x_1\theta_1 + \ldots + x_k\theta_k + x_1\theta_k + \ldots + x_k\theta_1$.

Thus we get

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{(1+e^{-y})} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix} + \frac{1}{(1+e^{y})} \begin{pmatrix} -\theta_k \\ \vdots \\ -\theta_1 \end{pmatrix} .$$

If we denote the components of $h_{\underline{x}}(\underline{\theta})$ by $h_{\underline{x}}^1(\underline{\theta}),\ldots,h_{\underline{x}}^k(\underline{\theta})$ respectively, then $h_{\underline{x}}^1(\underline{\theta}) \leq \ldots \leq h_{\underline{x}}^k(\underline{\theta})$ and so $A_{\underline{x}} = \{(d_1,\ldots,d_k): d_1 \leq \ldots \leq d_k\}$.

We can observe more reduction of $\mathcal{A}$ when $k = 2$. In this case

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{(1+e^{-y})} \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix} + \frac{1}{(1+e^{y})} \begin{pmatrix} -\theta_2 \\ -\theta_1 \end{pmatrix} ,$$

where $y = (x_1+x_2)(\theta_1+\theta_2)$. Now $h_{\underline{x}}^1(\underline{\theta}) + h_{\underline{x}}^2(\underline{\theta}) = (\theta_1+\theta_2)(e^{y}-e^{-y})$, which is positive, negative or zero according as (x_1+x_2) is positive, negative or zero. Therefore,

$$\begin{aligned} A_x &= \{ (d_1,d_2): d_1 \leq d_2 \text{ and } d_1+d_2 \geq 0 \}, \quad \text{if } x_1+x_2 > 0 \\ &= \{ (d_1,d_2): d_1 \leq d_2 \text{ and } d_1+d_2 \leq 0 \}, \quad \text{if } x_1+x_2 < 0 \\ &= \{ (d_1,d_2): d_1 \leq d_2 \text{ and } d_1+d_2 = 0 \}, \quad \text{if } x_1+x_2 = 0 . \end{aligned}$$

The maximum likelihood estimator for this problem is

$$\underset{\sim}{\delta}_{1/2}(\underset{\sim}{x}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \text{if} \quad x_1 \leq x_2$$

$$= \frac{1}{2}\begin{pmatrix} x_1+x_2 \\ x_1+x_2 \end{pmatrix}, \quad \text{if} \quad x_1 > x_2$$

(see Remark 3.3.4). It has been shown that $\underset{\sim}{\delta}_{1/2}(\underset{\sim}{x})$ is inadmissible, however no improvement has been obtained so far. Theorem 5.2.7 also does not help us in improving $\underset{\sim}{\delta}(\underset{\sim}{x})$ as it falls in the region $A_{\underset{\sim}{x}}$ for each $\underset{\sim}{x}$.

Example 5.3.6 (Simultaneous estimation of two normal means): Let X_1 and X_2 be independent normal random variables with means θ_1 and θ_2 and the common variance unity. We estimate $\underset{\sim}{\theta} = (\theta_1, \theta_2)$ with the loss function (5.3.6). The problem is invariant under the group G of all rotations in the plane, given by

$$G = G_1 \cup G_2, \quad \text{where}$$

$$G_1 = \left\{ g_a = \begin{bmatrix} \cos a & \sin a \\ -\sin a & \cos a \end{bmatrix}, \quad -\pi \leq a \leq \pi \right\}$$

and $$G_2 = \left\{ g_a^* = \begin{bmatrix} \cos a & \sin a \\ \sin a & -\cos a \end{bmatrix}, \quad -\pi < a < \pi \right\}.$$

It is well known that G is a compact group. The induced groups $\bar{G}$ and $\tilde{G}$ are defined in the same way as G. A G-equivariant estimator will be

$$d(\underset{\sim}{x}) = \underset{\sim}{x}\, \Psi(|\underset{\sim}{x}|).$$

We take ν to be uniform measure on a over the interval $[-\pi, \pi]$. Then we can write

$$\alpha(\underline{x}\mid\bar{g}_a(\underline{\theta})) = \frac{2\pi\, e^{u_1\cos a+u_2\sin a}}{\int_{-\pi}^{\pi}(e^{u_1\cos b+u_2\sin b}+e^{v_1\cos b+v_2\sin b})db},$$

and $$\alpha(\underline{x}\mid\bar{g}^*_a(\theta)) = \frac{2\pi\, e^{v_1\cos a+v_2\sin a}}{\int_{-\pi}^{\pi}(e^{u_1\cos b+u_2\sin b}+e^{v_1\cos b+v_2\sin b})db},$$

where $u_1 = x_1\theta_1+x_2\theta_2$, $u_2 = x_1\theta_2-x_2\theta_1$, $v_1 = x_1\theta_1-x_2\theta_2$ and $v_2 = x_1\theta_2+x_2\theta_1$.

Therefore

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{2\pi}\int_{-\pi}^{\pi}[\bar{g}_a(\underline{\theta})\ \alpha(\underline{x}\mid\bar{g}_a(\underline{\theta})) + \bar{g}^*_a(\underline{\theta})\ \alpha(\underline{x}\mid\bar{g}^*_a(\underline{\theta}))]\ da$$

$$= \begin{bmatrix} \dfrac{\int_{-\pi}^{\pi}(\theta_1\cos a+\theta_2\sin a)(e^{u_1\cos a+u_2\sin a}+e^{v_1\cos a+v_2\sin a})da}{\int_{-\pi}^{\pi}(e^{u_1\cos b+u_2\sin b}+e^{v_1\cos b+v_2\sin b})db} \\ \\ \dfrac{\int_{-\pi}^{\pi}(-\theta_1\sin a+\theta_2\cos a)(e^{u_1\cos a+u_2\sin a}-e^{v_1\cos a+v_2\sin a})da}{\int_{-\pi}^{\pi}(e^{u_1\cos b+u_2\sin b}+e^{v_1\cos b+v_2\sin b})db} \end{bmatrix} \tag{5.3.9}$$

Simplifying the right hand side of (5.3.9) and writing

$$I = \int_{0}^{\pi} e^{u_1\cos a}\cosh(u_2\sin a)da,$$

$$J = \int_{0}^{\pi} e^{v_1\cos a}\cosh(v_2\sin a)da,$$

$$I_1 = \int_{0}^{\pi}\cos a\ e^{u_1\cos a}\cosh(u_2\sin a)da,$$

$$I_2 = \int_{0}^{\pi}\sin a\ e^{u_1\cos a}\sinh(u_2\sin a)da,$$

$$J_1 = \int_0^{\pi} \cos a \; e^{v_1 \cos a} \cosh(v_2 \sin a) da ,$$

and $$J_2 = \int_0^{\pi} \sin a \; e^{v_1 \cos a} \sinh(v_2 \sin a) da ,$$

we get

$$h_{\underline{x}}(\underline{\theta}) = \frac{1}{(I+J)} \begin{bmatrix} \theta_1 I_1 + \theta_2 I_2 + \theta_1 J_1 + \theta_2 J_2 \\ -\theta_1 I_2 + \theta_2 I_1 + \theta_1 J_2 - \theta_2 J_1 \end{bmatrix} .$$

An integration by parts yields $I_1 = \frac{u_1}{u_2} I_2$ and $J_1 = \frac{v_1}{v_2} J_2$ and so

$$h_{\underline{x}}(\underline{\theta}) = \frac{(\theta_1^2 + \theta_2^2)}{(I + J)} \left(\frac{I_2}{u_2} + \frac{J_2}{v_2}\right) \underline{x} .$$

It can be easily verified that $\frac{I_2}{u_2}$ and $\frac{J_2}{v_2}$ are both nonnegative for all $\underline{\theta}$ and $\underline{x}$. Also I and J are nonnegative and so

$$A_{\underline{x}} = \{(d_1, d_2): d_i \text{ has the same sign as } x_i \text{ and } d_i = 0 \text{ whenever } x_i = 0,\ i = 1,2\} . \tag{5.3.10}$$

By Theorem 5.2.7 an admissible equivariant estimator lies in $A_{\underline{x}}$ and thus it should take values in the same quadrant as (x_1, x_2).

Remark 5.3.1: Let the parameter space in the above problem be restricted to $\Omega_o = \{(\theta_1, \theta_2): \theta_1^2 + \theta_2^2 \le r,\ r > 0\}$. In this case $A_{\underline{x}}$ is the $A_{\underline{x}}$ of (5.3.10) restricted to Ω_o.

CHAPTER - 6

Estimation of the Common Location Parameter of Two Populations with Different Scale Parameters

6.1 Introduction

Let $\underset{\sim}{X} = (X_1,\ldots,X_m)$ and $\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from two populations with densities $\frac{1}{\sigma_x} f(\frac{x-\theta}{\sigma_x})$ and $\frac{1}{\sigma_y} f(\frac{x-\theta}{\sigma_y})$ respectively, where σ_x and σ_y are unknown and possibly unequal. Let the density f be symmetric about zero. We are interested in the estimation of θ, the common location parameter. Available to us are unbiased estimators of θ based on the individual samples and we will like to improve at least one of them by a combination of the two. Hogg (1960b) suggested a method for combining the unbiased estimators from the individual samples, so that the resulting estimator is also unbiased. Let Ψ_x and γ_x be odd location-scale and even location free-scale statistics (these terms are defined in Section 6.2) respectively, based on $\underset{\sim}{X}$. Hogg (1960a) proved that Ψ_x and γ_x are uncorrelated. As a consequence of this result, Hogg (1960b) showed that the conditional expectation of Ψ_x given γ_x equals θ. Thus, if Ψ_x and Ψ_y are odd location-scale estimators of θ based on $\underset{\sim}{X}$ and $\underset{\sim}{Y}$ respectively, an unbiased estimator which combines Ψ_x and Ψ_y is $W\Psi_x + (1-W)\Psi_y$, where W is a function of even location free-scale statistics γ_x and γ_y.

Cohen (1976) discussed some situations where the problem of estimating a common location parameter arises. For example, two laboratories might be evaluating the same product. It may be

quite appropriate to assume that the locations of the measured aspect of the product are the same, but the scales differ because of laboratory techniques or facilities. The assumption on the distribution of the measured quantity could very well be a particular location-scale family, or perhaps a finite set of location scale families.

Cohen considered improving Ψ_x or Ψ_y when the loss function is squared error. He suggested using an estimator δ_a and showed that under a simple moment condition, δ_a improves Ψ_x for $0 < a \leq a^*(m,n)$, where $a^*(m,n)$ is a constant depending on m and n only. He evaluated $a^*(m,n)$ for specific values of m and n when the underlying distribution is uniform.

Bhattacharyya (1981) showed that the above class of estimators δ_a dominating Ψ_x can be enlarged in the sense that there is an $A(m,n) > a^*(m,n)$ such that δ_a for $0 < a \leq A(m,n)$ is better than Ψ_x. The constant $A(m,n)$ is easier to calculate than $a^*(m,n)$. Bhattacharyya tabulated values of $A(m,n)$ for specific choices of m and n in case of uniform densities. It was observed that Cohen's bound $a^*(m,n)$ does not exceed one for both the pairs (m,n) and (n,m) simultaneously for $m = 2(1)15(5)50$ and $n = 6(1)15(5)50$. However $A(m,n)$ exceeds one for $m \geq 25$ when $m = n$ and for both the pairs (m,n) and (n,m) for $n \geq 35$ whenever $m \leq n+5$. As will be seen later in Section 6.2 if the bound exceeds one for both the pairs (m,n) and (n,m), the implication is that the estimator δ_a with $a = 1$ dominates both Ψ_x and Ψ_y.

In Section 6.2 of this chapter, we give the general set up and the necessary notation. Also the results of Cohen and

Bhattacharyya are stated. In Section 6.3, we develop sufficient conditions under which the bounds, different from those of Cohen and Bhattacharyya, can be obtained. These conditions and the resulting bounds are given in Theorems 6.3.1 and 6.3.2. The bound obtained from Theorem 6.3.1 is simpler to evaluate than $a^*(m,n)$ and $A(m,n)$. However, it does not always improve $A(m,n)$. The bound obtained from Theorem 6.3.2 is always larger than $A(m,n)$ and also simpler to evaluate. For uniform distribution, we observe substantial improvements over $A(m,n)$. If we denote the improving bound by $B_4(m,n)$, it is seen to exceed one for both the pairs (m,n) and (n,m) for $m \geq 20$, if $m = n$ and for $n \geq 32$, if $m \leq n+5$ compared to $A(m,n)$ with this property for $m \geq 25$, if $m = n$ and for $n \geq 35$, if $m \leq n+5$. Thus, we are able to prove the dominance of δ_a, over both Ψ_x and Ψ_y for the values of m, n, for which Bhattacharyya shows dominance only over Ψ_x.

In Section 6.4, we consider another estimator δ_c^* and obtain conditions under which it dominates one or both of Ψ_x and Ψ_y. The conditions are stated in Theorems 6.4.1, 6.4.2 and 6.4.3. An advantage in considering δ_c^* is that we get a class of estimators including δ_1, which dominate both Ψ_x and Ψ_y whereas in Section 6.3 we could have at most one estimator, namely δ_1, dominating both Ψ_x and Ψ_y.

Akai (1982) has also obtained a sufficient condition for δ_c^* to dominate Ψ_x and Ψ_y. His result is stronger than the one given in Theorem 6.4.2(i).

In Section 6.5, the results of Section 6.3 are applied to the uniform distribution and estimators improving Ψ_x and/or Ψ_y are obtained.

In Section 6.6, we consider estimation of the common mean μ of two normal populations with unknown and unequal variances σ_1^2 and σ_2^2. A commonly used estimator of μ is the Graybill-Deal estimator (1959) given by

$$\hat{\mu}_{GD} = \frac{S_2\bar{X} + S_1\bar{Y}}{S_1 + S_2},$$

where $\bar{X}$, $\bar{Y}$, S_1 and S_2 are the sample means and sample variances. Sinha and Mouqadem (1982) considered a class $\mathcal{C}$ of unbiased estimators of μ;

$$\mathcal{C} = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\ \Psi(S_1, S_2, (\bar{Y}-\bar{X})^2), \quad 0 \leq \Psi \leq 1\}.$$

They then obtained an expression for the risk of an estimator $\delta \in \mathcal{C}$ and studied admissibility of $\hat{\mu}_{GD}$ in various subclasses of $\mathcal{C}$. We consider improving $\hat{\mu}_{GD}$ in a subclass $\mathcal{C}_1 = \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\ \Psi(S_1,S_2),\ 0 \leq \Psi \leq 1\}$ of $\mathcal{C}$ and derive a differential inequality using an identity due to Hudson (1978). A solution to this inequality will provide an improvement over $\hat{\mu}_{GD}$. However, the question of the existence of a solution still remains unresolved.

In Section 6.7, we consider the problem of estimating the common mean of a bivariate normal population. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a random sample from a $N_2\left(\begin{pmatrix}\mu\\ \mu\end{pmatrix}, \begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}\right)$ distribution with ρ unknown. Krishnamoorthy and Rohatgi (1988a) considered estimators

$$\hat{\mu}_3(c) = \frac{1}{2}\left(\bar{V} - c\bar{U}\ \frac{S^*_{UV}}{\sum_{i=1}^{n} U_i^2}\right), \qquad c \text{ a real constant,}$$

where $U_i = X_i - Y_i$, $V_i = X_i + Y_i$, $i = 1, 2, \ldots, n$;

$$\bar{U} = \frac{1}{n}\sum_{i=1}^{n} U_i, \quad \bar{V} = \frac{1}{n}\sum_{i=1}^{n} V_i, \quad S^*_{UV} = \sum_{i=1}^{n} (U_i - \bar{U})(V_i - \bar{V}),$$

In particular, they considered $\hat{\mu}_3 = \hat{\mu}_3(1)$ and obtained the region of the parameter space where $\hat{\mu}_3$ dominates the MLE $\hat{\mu}_2$. We obtain a c^* such that $\hat{\mu}_3(c^*)$ has the largest region of dominance over $\hat{\mu}_2$ among estimators of the form $\hat{\mu}_3(c)$. Also we show that a reasonable choice of c is $\frac{n}{n-1}$ and $\hat{\mu}_3(\frac{n}{n-1})$ has a larger region of dominance than $\hat{\mu}_3$. Further we obtain minimal essentially complete subclass of $\{\hat{\mu}_3(c),\ c \text{ real}\}$ and compare numerically the risk functions of $\hat{\mu}_2$, $\hat{\mu}_3$, $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(c^*)$. Once again, the choice is between $\hat{\mu}_3(c^*)$ and $\hat{\mu}_3(\frac{n}{n-1})$.

6.2 Preliminaries

Let $\underline{X} = (X_1, \ldots, X_m)$ be a random sample of size m from a population with the density $\frac{1}{\sigma_x} f(\frac{x-\theta}{\sigma_x})$, $\sigma_x > 0$. Let $\underline{Y} = (Y_1, \ldots, Y_n)$ be a random sample, independent of $\underline{X}$, from a population with the density $\frac{1}{\sigma_y} f(\frac{x-\theta}{\sigma_y})$, $\sigma_y > 0$. The function f is assumed to be symmetric about zero. The problem is one of estimating the common location parameter θ when the scale parameters σ_x and σ_y are unknown and unequal. The loss function is taken to be squared error. However, attention is restricted to the unbiased estimators of θ and so the risk of any estimator is simply its variance.

Further, we introduce invariance in the above problem. Let G be the affine group of transformations operating on the sample space, that is,

$$G = \{g: g_{a,b}(x) = ax+b,\ a \text{ and } b \text{ real}\}.$$

We have then under $g_{a,b}$:

$$X_i \to aX_i+b\ ,\quad i = 1,\ldots,m\ ,$$
$$Y_j \to aY_j+b\ ,\quad j = 1,\ldots,n\ ,$$
$$\theta \to a\theta+b\ ,$$
$$\sigma_x \to |a|\sigma_x \text{ and}$$
$$\sigma_y \to |a|\sigma_y\ .$$

The estimation problem is invariant when the loss function is

$$L(\theta,a) = (\theta-a)^2/\sigma_x^2 \tag{6.2.1}$$

However, there is no change in the results, if instead of (6.2.1) we proceed with the squared error loss function.

Let Ψ_x be a G-equivariant estimator of θ. Thus, Ψ_x satisfies

$$\Psi_x(ax_1+b,\ldots,ax_m+b) = a\,\Psi_x(x_1,\ldots,x_m)+b \tag{6.2.2}$$

for every real a and b. The statistics Ψ_x satisfying (6.2.2) are also called odd location-scale statistics. Let γ_x be even location free-scale invariant estimator of σ_x. That is, γ_x satisfies

$$\gamma_x(ax_1+b,\ldots,ax_m+b) = |a|\,\gamma_x(x_1,\ldots,x_m) \tag{6.2.3}$$

for every real a and b.

Analogously, statistics Ψ_y and γ_y are defined for the Y-sample.

We notice that the random variables $T_x = (\Psi_x-\theta)/\sigma_x$, $T_y = (\Psi_y-\theta)/\sigma_y$, $S_x = \gamma_x/\sigma_x$ and $S_y = \gamma_y/\sigma_y$ have distributions which do not depend on any parameters.

Further, define $U = S_y/S_x$, $V = U^2$, $Z = \gamma_y^2/\gamma_x^2$, $p = \sigma_y^2/(\sigma_x^2+\sigma_y^2)$, $k(v) = \min(1,1/v)$, $h(v) = \max(1,1/v)$ and $g(T_x^2,T_y^2) = \max(T_x^2,T_y^2)$.

Cohen (1976) suggested estimator δ_a for θ defined by

$$\delta_a = \left(1 - \frac{a}{1+Z}\right) \Psi_x + \frac{a}{1+Z} \Psi_y , \tag{6.2.4}$$

where a is a positive real. He proved the following result.

Theorem 6.2.1 (Cohen (1976)): Assume $E g(T_x^2, T_y^2) h^2(V) < \infty$. Let f be symmetric about 0, and let

$$a^*(m,n) = \frac{2E\, T_x^2\, k(V)}{E\, g(T_x^2, T_y^2) h^2(V)} . \tag{6.2.5}$$

Then for all a such that $0 < a \leq a^*(m,n)$, the estimator δ_a is unbiased and improves upon Ψ_x when the loss function is squared error.

Remark 6.2.1: If $a^*(m,n) \geq 1$ for both pairs (m,n) and (n,m) then symmetry implies that δ_1 is better than both Ψ_x and Ψ_y.

The constant $a^*(m,n)$ is obtained as a lower bound for $M = \inf_{0 \leq p < 1} R(p)$ where R(p) is given by

$$R(p) = \frac{2E\, W\, T_x^2}{E W^2 (p T_y^2 + (1-p) T_x^2)} , \tag{6.2.6}$$

and

$$W = \frac{1}{1 - p + pV} .$$

Bhattacharyya (1981) obtained a sharper lower bound A(m,n) for M and proved the following result.

Theorem 6.2.2 (Bhattacharyya (1981)): Assume $E(T_x^2 V^{-2})$ and $E T_y^2 h^2(V)$ are finite. Let f be symmetric about 0, and let

$$A(m,n) = 2 \min(1, A_1(m,n), A_2(m,n)) , \tag{6.2.7}$$

where

$$A_1(m,n) = E(T_x^2 V^{-1})/E(T_x^2 V^{-2}) , \qquad (6.2.8)$$

and

$$A_2(m,n) = E\ T_x^2\ k(V)/E\ T_y^2 h^2(V) . \qquad (6.2.9)$$

Then for all a such that $0 < a \leq A(m,n)$, the estimator δ_a dominates Ψ_x.

Remark 6.2.2: As earlier, δ_1 will be better than both Ψ_x and Ψ_y, if $A(m,n) \geq 1$ for both the pairs (m,n) and (n,m).

Remark 6.2.3: The difference between Cohen (1976) and Bhattacharyya (1981) approaches is that Cohen finds lower and upper bounds for the integrands whereas Bhattacharyya finds these bounds for the integral itself. If a random vector $(z_1,\ldots,z_k)$ has $(\theta_1,\ldots,\theta_k)$ as its mean vector, then $Ez_{min} < \theta_{min}$ and $Ez_{max} > \theta_{max}$. This fact explains why Bhattacharyya's bound is larger than Cohen's bound.

6.3 Improving upon the Bounds of Cohen and Bhattacharyya

The estimator δ_a will be better than Ψ_x iff

$$0 < a \leq M = \inf_{0 \leq p < 1} R(p) . \qquad (6.3.1)$$

Consider

$$R(p) = \frac{2E\ W\ T_x^2}{EW^2((1-p)\ T_x^2 + p\ T_y^2)} . \qquad (6.3.2)$$

Clearly

$$R(p) \geq 2 \min(M_1(p), M_2(p)) , \qquad (6.3.3)$$

where $M_1(p) = E\ W\ T_x^2/E\ W^2 T_x^2$

and $M_2(p) = E\ W\ T_x^2/E\ W^2 T_y^2$.

In Bhattacharyya's bound $A(m,n)$, the terms $\min(1,A_1(m,n))$ and $A_2(m,n)$ were obtained as lower bounds for $M_1(p)$ and $M_2(p)$ respectively. However, $A_2(m,n)$ as a bound for $M_2(p)$ can be improved. We give below two other lower bounds for $M_2(p)$, one of which is better than $A_2(m,n)$, the other one though not always better is much simpler to calculate.

Rewrite $M_2(p)$ as $EWg_1(U)/EW^2g_2(U)$, where $g_1(u) = E(T_x^2 \mid U=u)$ and $g_2(u) = E(T_y^2 \mid U=u)$. If we denote by $h_1(u)$ the density of U, then

$$M_2(p) = \frac{\int w\, g_1(u)\, h_1(u)\, du}{\int w^2\, g_2(u)\, h_1(u) du}$$

$$= E_1(1-p+pU^2)\,\frac{g_1(U)}{g_2(U)}$$

$$\geq \min\left(E_1\,\frac{g_1(U)}{g_2(U)},\; E_1\,\frac{U^2 g_1(U)}{g_2(U)}\right),$$

where E_1 denotes the expectation with respect to the density $h_1(u,p) = \dfrac{w^2\, g_2(u)\, h_1(u)}{\int w^2\, g_2(u)\, h_1(u)du}$. The fact that the density $h_1(u,p)$ has a monotone likelihood ratio (MLR) in $(p,-u)$, that is, for $p_1 > p_2$, $u_1 < u_2$;

$$h_1(u_1,p_1)\, h_1(u_2,p_2) - h_1(u_1,p_2)\, h_1(u_2,p_1) \geq 0,$$

then proves the following theorem (see Lehmann (1959, p. 74) for the related property of the densities having MLR).

<u>Theorem 6.3.1</u>: (i) Let $E(T_x^2\, V^{-2})$ and $E(T_y^2\, V^{-2})$ be finite and let f be symmetric about 0. Also let $\dfrac{g_1(u)}{g_2(u)}$ be an increasing function of u and

$$B_1(m,n) = 2\min(1,A_1(m,n),A_3(m,n)), \qquad (6.3.4)$$

where $A_1(m,n)$ is as in (6.2.8) and

$$A_3(m,n) = \min\left(\frac{E(T_x^2\, V^{-2})}{E(T_y^2\, V^{-2})}, \frac{E(T_x^2\, V^{-1})}{E(T_y^2\, V^{-2})}\right). \qquad (6.3.5)$$

Then for all a such that $0 < a \leq B_1(m,n)$, the estimator δ_a dominates Ψ_x.

(ii) Let $E(T_x^2\, V^{-2})$, $E(T_y^2)$ and $E(T_y^2\, V^{-2})$ be all finite and let f be symmetric about 0. Also let $\frac{g_1(u)}{g_2(u)}$ be a decreasing function of u and let $\frac{u^2 g_1(u)}{g_2(u)}$ be an increasing function of u and

$$B_2(m,n) = 2\min(1, A_1(m,n), A_4(m,n)), \qquad (6.3.6)$$

where

$$A_4(m,n) = \min\left(\frac{E(T_x^2)}{E(T_y^2)}, \frac{E(T_x^2\, V^{-1})}{E(T_y^2\, V^{-2})}\right). \qquad (6.3.7)$$

Then for all a such that $0 < a \leq B_2(m,n)$, the estimator δ_a dominates Ψ_x.

(iii) Let $E(T_y^2)$ and $E(T_x^2\, V^{-2})$ be finite and let f be symmetric about 0. Also let $u^2\frac{g_1(u)}{g_2(u)}$ be a decreasing function of u and

$$B_3(m,n) = 2\min(1, A_1(m,n), A_5(m,n)), \qquad (6.3.8)$$

where

$$A_5(m,n) = \min\left(\frac{E(T_x^2)}{E(T_y^2)}, \frac{E(T_x^2 V)}{E(T_y^2)}\right). \qquad (6.3.9)$$

Then for all a such that $0 < a \leq B_3(m,n)$, the estimator δ_a is better than Ψ_x.

<u>Remark 6.3.1</u>: In the expression for $A_3(m,n)$ the second component is always larger than $A_2(m,n)$ but the first component is sometimes smaller than $A_2(m,n)$. Thus $B_1(m,n)$ is not always better than Bhattacharyya's bound $A(m,n)$.

<u>Remark 6.3.2</u>: $A_4(m,n)$ is always larger than $A_2(m,n)$ and so $B_2(m,n)$ improves $A(m,n)$.

<u>Remark 6.3.3</u>: In $A_5(m,n)$ the first component is larger than $A_2(m,n)$ but the second is not always so with the consequence that $B_3(m,n)$ is not always better than $A(m,n)$.

<u>Remark 6.3.4</u>: The constants $B_i(m,n)$, $i = 1,2,3$ are relatively simpler to compute than $A(m,n)$.

<u>Remark 6.3.5</u>: Using the MLR argument as applied in the proof of the above theorem, one can get an alternative proof of Bhattacharyya's claim $M_1(p) \geq 2 \min(1, A_1(m,n))$.

To obtain another lower bound on $M_2(p)$, we now consider

$$\frac{1}{M_2(p)} = \frac{\int w^2 \, g_2(u) \, h_1(u) \, du}{\int w \, g_1(u) \, h_1(u) \, du}$$

$$= E_2 \, \frac{W \, g_2(U)}{g_1(U)}$$

$$\leq E_2 \, \frac{g_2(U)}{g_1(U)} \, h(U^2) \, ,$$

where E_2 is the expectation with respect to the density

$$h_2(u,p) = \frac{w \, g_1(u) \, h_1(u)}{\int w \, g_1(u) \, h_1(u) \, du} \, .$$

Once again using the fact that the density $h_2(u,p)$ has MLR in $(p,-u)$ the following result can be proved.

<u>Theorem 6.3.2</u>: (i) Let $E(T_x^2 \, V^{-2})$ and $E(T_y^2 \, h(V) \, V^{-1})$ be finite and let f be symmetric about 0. Also let $\frac{g_2(u)}{g_1(u)} \, h(u^2)$ be a decreasing function of u and

$$B_4(m,n) = 2\min(1, A_1(m,n), A_6(m,n)) , \tag{6.3.10}$$

where

$$A_6(m,n) = E(T_x^2 V^{-1})/E(T_y^2 h(V) V^{-1}) . \tag{6.3.11}$$

Then for all a such that $0 < a \leq B_4(m,n)$, the estimator δ_a is better than Ψ_x.

(ii) Let $E(T_x^2 V^{-2})$ and $E(T_y^2 h(V))$ be finite and let f be symmetric about 0. Also let $\frac{g_2(u)}{g_1(u)} h(u^2)$ be an increasing function of u and

$$B_5(m,n) = 2\min(1, A_1(m,n), A_7(m,n)) , \tag{6.3.12}$$

and

$$A_7(m,n) = E(T_x^2)/E(T_y^2 h(V)) . \tag{6.3.13}$$

Then for all a such that $0 < a \leq B_5(m,n)$, the estimator δ_a is better than Ψ_x.

<u>Remark 6.3.6</u>: The constants $B_4(m,n)$ and $B_5(m,n)$ are never smaller than $A(m,n)$ and so in both the cases Bhattacharyya's bound is improved. As will be seen in Section 6.5, for the uniform distribution, there is substantial improvement over $A(m,n)$.

<u>Remark 6.3.7</u>: The conditions for simultaneous improvement over both the estimators Ψ_x and Ψ_y are similar to the one in Remark 6.2.1 following Theorem 6.2.1. Thus, the estimator δ_1 improves both Ψ_x and Ψ_y if the bounds obtained in Theorems 6.3.1 and 6.3.2 are not less than 1 for both the pairs (m,n) and (n,m).

6.4 A Class of Estimators Improving upon the Estimators from Both the Samples

Shinozaki (1978) considered estimators of the form

$$\delta_c^* = W_1\Psi_x + W_2\Psi_y , \tag{6.4.1}$$

where

$$W_1 = \frac{c\,\gamma_x^{-2}}{c\,\gamma_x^{-2} + \gamma_y^{-2}} \quad \text{and} \quad W_2 = 1-W_1 .$$

When the populations under consideration are normal, he obtained conditions on c such that δ_c^* improves both Ψ_x and Ψ_y. Similar results were obtained later by Bhattacharyya (1984) and Kubokawa (1987c) but only for the case of normal populations. We consider this problem in the location-scale set up introduced in Section 6.2.

Using the results of Hogg (1960b) it is easily seen that δ_c^* is unbiased and the condition $\mathrm{Var}(\delta_c^*) \leq \mathrm{Var}(\Psi_x)$ for all $(\theta, \sigma_x^2, \sigma_y^2)$ is equivalent to $M(c) \geq 1$ where $M(c) = \inf\limits_{0 \leq p < 1} R(p,c)$, $R(p,c) = \dfrac{2EW^* T_x^2}{EW^{*2}((1-p)T_x^2 + pT_y^2)}$ and $W^* = 1/(1-p+pcV)$. As M(c) is, in general, difficult to calculate, we find some lower bounds for M(c). The methods used are similar to the ones in Section 6.3.

Now

$$\frac{1}{2} R(p,c) \geq \min(M_1(p,c), M_2(p,c)) , \tag{6.4.2}$$

where $M_1(p,c) = E(W^* T_x^2)/E(W^{*2} T_x^2)$, (6.4.3)

and $M_2(p,c) = E(W^* T_x^2)/E(W^{*2} T_y^2)$. (6.4.4)

If we write $g^*(v)$ for $E(T_x^2 \mid V=v)$, $M_1(p,c)$ is

$$= \frac{EW^* g^*(V)}{EW^{*2} g^*(V)}$$

$$= \frac{\int w^* g^*(v)\, h^*(v)\, dv}{\int w^{*2} g^*(v)\, h^*(v)\, dv}$$

$$= E^* \frac{1}{W^*} \geq \min(1, cE^*(V)) .$$

where in the above expressions, $h^*(v)$ denotes the probability density of V and E^* is the expectation with respect to the density

$$h^*(v,p) = \frac{w^{*2} g^*(v) h^*(v)}{\int w^{*2} g^*(v) h^*(v) dv} .$$

Since $h^*(v,p)$ has MLR in $(p,-v)$ we have $M_1(p,c) \geq \min(1, cA_1(m,n))$, where $A_1(m,n)$ is as defined in (6.2.8). Also, we have $k(cV) \leq W^* \leq h(cV)$ so that

$$M_2(p,c) \geq A_2(c,m,n) ,$$

where $$A_2(c,m,n) = E\{T_x^2 k(cV)\}/E\{T_y^2 h^2(cV)\} . \tag{6.4.5}$$

We have thus the following result.

Theorem 6.4.1: Let $E(T_x^2 V^{-2})$ be finite and let f be symmetric about 0. Then the estimator δ_c^* is better than Ψ_x for all c such that $ET_y^2 h^2(cV) < \infty$ and

$$2 \min(cA_1(m,n), A_2(c,m,n)) \geq 1 . \tag{6.4.6}$$

Remark 6.4.1: By symmetry, we can show that δ_c^* is better than Ψ_y, if $E(T_y^2 V^2)$ is finite and c is such that $ET_x^2 h^2(1/cV) < \infty$ and

$$2 \min(\tfrac{1}{c} A_1(n,m), A_2(\tfrac{1}{c}, n,m)) \geq 1 . \tag{6.4.7}$$

Remark 6.4.2: Combining (6.4.6) and (6.4.7) we get the class of estimators δ_c^* dominating both Ψ_x and Ψ_y.

Next, we obtain some other lower bounds for $M_2(p,c)$. In the notation of Section 6.3, we can write

$$M_2(p,c) = \frac{EW^* g_1(U)}{EW^{*2} g_2(U)}$$

$$= \frac{\int w^* g_1(u)\, h_1(u)\, du}{\int w^{*2} g_2(u)\, h_1(u)\, du}$$

$$= E_1^* \frac{g_1(U)}{W^* g_2(U)}$$

$$\geq \min(E_1^* \frac{g_1(U)}{g_2(U)}, \quad cE_1^* \frac{U^2 g_1(U)}{g_2(U)}) ,$$

where E_1^* is the expectation with respect to the density $h_1^*(u,p) = \frac{w^{*2} g_2(u)\, h_1(u)}{\int w^{*2} g_2(u)\, h_1(u)\, du}$, which has MLR in $(p,-u)$.

Define

$$P_1(m,n) = \frac{E(T_x^2 V^{-2})}{E(T_y^2 V^{-2})}, \qquad P_2(m,n) = \frac{E(T_x^2 V^{-1})}{E(T_y^2 V^{-2})},$$

$$P_3(m,n) = \frac{E(T_x^2)}{E(T_y^2)} \quad \text{and} \quad P_4(m,n) = \frac{E(T_x^2 V)}{E(T_y^2)} . \tag{6.4.8}$$

Then we have the following theorem.

<u>Theorem 6.4.2</u>: Let f be symmetric about 0 and either of the conditions (i), (ii) and (iii) stated below hold.

(i) $E(T_x^2 V^{-2})$ and $E(T_y^2 V^{-2})$ are finite, $\frac{g_1(u)}{g_2(u)}$ is increasing in u and

$$2 \min(cA_1(m,n), P_1(m,n), cP_2(m,n)) \geq 1 . \tag{6.4.9}$$

(ii) $E(T_x^2 V^{-2})$, $E(T_y^2)$ and $E(T_y^2 V^{-2})$ are finite, $\frac{g_1(u)}{g_2(u)}$ is decreasing in u, $\frac{u^2 g_1(u)}{g_2(u)}$ is increasing in u and

$$2 \min(cA_1(m,n), P_3(m,n), cP_2(m,n)) \geq 1 . \tag{6.4.10}$$

(iii) $E(T_x^2 V^{-2})$ and $E(T_y^2)$ are finite, $\frac{u^2 g_1(u)}{g_2(u)}$ is decreasing in u and

$$2\min(cA_1(m,n), P_3(m,n), cP_4(m,n)) \geq 1 . \tag{6.4.11}$$

Then δ_c^* improves Ψ_x.

Remark 6.4.3: By symmetry we again get the conditions for δ_c^* to dominate Ψ_y and combining these with (i), (ii) and (iii), we get the class of estimators dominating both Ψ_x and Ψ_y.

Remark 6.4.4: Akai (1982, Cor. 2.3) has given a sufficient condition for δ_c^* to dominate Ψ_x. He proved that if

$$E(T_x^2V^{-2})/E(T_y^2V^{-2}) \leq E(T_x^2V^{-1})/E(T_y^2V^{-1}) , \tag{6.4.12}$$

δ_c^* improves Ψ_x for $c \geq 1/2P_2(m,n)$.

It can be shown that Akai's condition is weaker than condition (i) of Theorem 6.4.2. The proof uses the following integral inequality (see Bhattacharyya (1984) for details):

Let Z_1, Z_2 and Z_3 be functions of a random variable T; Z_2 be positive with a finite expectation and Z_1/Z_2 and Z_3 be monotonic in opposite directions. Then

$$E(Z_1Z_3)/E(Z_2Z_3) \leq EZ_1/EZ_2 \tag{6.4.13}$$

Choosing $T = U$, $Z_1 = U^{-2}g_1(U)$, $Z_2 = U^{-2}g_2(U)$ and $Z_3 = U^{-2}$, we see that condition (i) of Theorem 6.4.2 implies (6.4.12).

However, Theorem 6.4.2 also applies to situations where the condition (6.4.12) is not satisfied. Note that if $\frac{g_1(u)}{g_2(u)}$ is decreasing u, then

$$E(T_x^2V^{-2})/E(T_y^2V^{-2}) \geq E(T_x^2V^{-1})/E(T_y^2V^{-1})$$

and Akai's result is not applicable.

Once again consider $M_2(p,c)$:

$$\frac{1}{M_2(p,c)} = \frac{E\ W^{*2}\ g_2(U)}{E\ W^*\ g_1(U)}$$

$$= \frac{\int w^{*2}\ g_2(u)\ h_1(u)\ du}{\int w^*\ g_1(u)\ h_1(u)\ du}$$

$$= E_2^*\ W^*\ \frac{g_2(U)}{g_1(U)}$$

$$\geq E_2^*\ \frac{g_2(U)}{g_1(U)}\ h(cU^2)\ ,$$

where E_2^* is the expectation with respect to the density $h_2^*(u,p) = \dfrac{w^*\ g_1(u)\ h_1(u)}{\int w^*\ g_1(u)\ h_1(u)\ du}$ which has MLR in $(p,-u)$.

Define

$$P_5(c,m,n) = \frac{ET_x^2}{E\ T_y^2\ h(cV)}\ ,$$

$$\text{and} \qquad P_6(c,m,n) = \frac{E(T_x^2V^{-1})}{E(T_y^2\ h(cV)\,V^{-1})}\ . \tag{6.4.14}$$

Then we have the following result.

<u>Theorem 6.4.3</u>: Let f be symmetric about 0 and either of the conditions (i) and (ii) stated below be satisfied

(i) $E(T_x^2V^{-2})$ and $E(T_y^2h(cV))$ are finite, $\dfrac{g_2(u)}{g_1(u)}\ h(u^2)$ is increasing in u and

$$2\ \min(cA_1(m,n),P_5(c,m,n)) \geq 1\ , \tag{6.4.15}$$

(ii) $E(T_x^2V^{-2})$ and $E(T_y^2h(cV)\,V^{-1})$ are finite, $\dfrac{g_2(u)}{g_1(u)}\ h(u^2)$ is decreasing in u and

$$2\ \min(cA_1(m,n),P_6(c,m,n)) \geq 1\ . \tag{6.4.16}$$

Then δ_c^* dominates Ψ_x.

Remark 6.4.5: Symmetry considerations once again give conditions for simultaneous improvement of Ψ_x and Ψ_y. However, the class of such estimators is difficult to obtain here.

Remark 6.4.6: Let condition (i) of Theorem 6.4.2 be applicable, that is, $\frac{g_1(u)}{g_2(u)}$ be increasing in u. Then δ_c^* improves Ψ_x, if $2P_1(m,n) \geq 1$ and $c \geq 1/2\min(A_1(m,n), P_2(m,n))$. In this case condition (ii) of Theorem 6.4.3 also applies and we conclude that δ_c^* improves Ψ_x, if $c \geq 1/2A_1(m,n)$ and $2P_6(c,m,n) \geq 1$. Notice that $P_6(c,m,n) \leq cP_2(m,n)$ and so whenever $2P_1(m,n) \geq 1$, the class of estimators dominating Ψ_x obtained by Theorem 6.4.2 contains that obtained by Theorem 6.4.3. However, if $2P_1(m,n) < 1$, Theorem 6.4.2 fails to provide an estimator δ_c^* improving Ψ_x whereas Theorem 6.4.3 still gives such estimators.

Remark 6.4.7: We also have $A_2(c,m,n) \leq P_6(c,m,n)$ for each c, m and n. This means that the class of estimators δ_c^* dominating $\Psi_x(\Psi_y)$ obtained by Theorem 6.4.1 is smaller than that obtained by Theorem 6.4.3 provided condition (ii) of Theorem 6.4.3 holds. Once again, it may happen that Theorem 6.4.3 is not applicable and we use Theorem 6.4.1 to get δ_c^* dominating $\Psi_x(\Psi_y)$.

6.5 The Uniform Distribution

In this section we apply the results of Section 6.3 to the case when we have two populations having uniform densities which are symmetric about the common location parameter θ and have unknown and unequal scale parameters σ_x and σ_y. The bounds $B_1(m,n)$ and $B_4(m,n)$ of Theorems 6.3.1 and 6.3.2 are evaluated and

the larger of the two is tabulated and compared with Bhattacharyya' bound $A(m,n)$ for specific values of m and n.

Let $f(x) = 1, -\frac{1}{2} \le x \le \frac{1}{2}$. Also let Ψ_x and γ_x of Section 6.2 be $\frac{1}{2}(X_{(m)} + X_{(1)})$ and $X_{(m)} - X_{(1)}$ respectively, where $X_{(1)},\ldots,X_{(m)}$ are the order statistics of the first sample. Similarly define Ψ_y and γ_y for the second sample. The variables T_x, T_y, S_x, S_y, U and V are the same as in Section 6.2. For convenience, take $L_x = 2T_x$ and $L_y = 2T_y$. The joint probability density of L_x and S_x, as given by Cohen (1976), is

$$p_{L_x,S_x}(l_x, s_x) = \tfrac{1}{2}m(m-1)\; s_x^{m-2}, \quad -1+s_x < l_x < 1-s_x; \quad 0 < s_x < 1 .$$

The conditional probability density of L_x given $U = u$ is

$$\frac{1}{2}\,\frac{j\{(j-1)(1+l_x)^{j-2} - u(j-2)(1+l_x)^{j-1}\}}{\{j-u(j-2)\}}, \quad \text{if } -1<l_x\le 0,\ 0<u\le 1 ;$$

$$\frac{1}{2}\,\frac{j\{(j-1)(1-l_x)^{j-2} - u(j-2)(1-l_x)^{j-1}\}}{\{j-u(j-2)\}}, \quad \text{if } 0<l_x<1,\ 0<u\le 1 ;$$

$$\frac{1}{2}\,\frac{ju^{j-1}\{(j-1)(1+l_x)^{j-2} - u(j-2)(1+l_x)^{j-1}\}}{\{ju-(j-2)\}}, \quad \text{if } -1<l_x\le\frac{1}{u}-1,\ u>1 ;$$

$$\frac{1}{2}\,\frac{ju}{\{ju-(j-2)\}}, \quad \text{if } \frac{1}{u}-1 < l_x \le 1-\frac{1}{u}, \quad u>1 ; \quad \text{and}$$

$$\frac{1}{2}\,\frac{ju^{j-1}\{(j-1)(1-l_x)^{j-2} - u(j-2)(1-l_x)^{j-1}\}}{\{ju-(j-2)\}}, \quad \text{if } 1-\frac{1}{u}<l_x<1,\ u>1,$$

where $j = m+n$.

It can be seen that $g_1(u) = E(T_x^2 \mid U=u)$ is

$$\frac{1}{2}\,\frac{\{(j+2) - u(j-2)\}}{(j+1)(j+2)\{j-u(j-2)\}} = \alpha(u), \text{ say, if } 0 < u \le 1 \text{ and}$$

$$\frac{1}{2}\,\frac{j}{\{ju-(j-2)\}}\left[\frac{1}{j} + \frac{1-2j}{j(j+1)u} + \frac{j}{(j+1)(j+2)u^2} + \frac{u}{6}\left(1-\frac{1}{u}\right)^3\right]$$

$$= \beta(u), \text{ say, if } u > 1 .$$

Now by symmetry

$$g_2(u) = E(T_y^2 \mid U=u) = E(T_y^2 \mid \tfrac{1}{U} = \tfrac{1}{u}) = g_1(\tfrac{1}{u}) \ .$$

Thus,

$$g_2(u) = \begin{cases} \alpha(\frac{1}{u}), & \text{if } u \geq 1 \text{ and} \\ \beta(\frac{1}{u}), & \text{if } 0 < u < 1 \end{cases}$$

and so

$$\frac{g_1(u)}{g_2(u)} = \begin{cases} \frac{\alpha(u)}{\beta(1/u)}, & \text{if } 0 < u \leq 1 \\ \frac{\beta(u)}{\alpha(1/u)}, & \text{if } u > 1 \ . \end{cases}$$

It can be shown that the numerator in the expression of $\frac{d}{du}\left(\frac{\alpha(u)}{\beta(1/u)}\right)$ is positive for all $u \in (0,1]$ and so $\frac{\alpha(u)}{\beta(1/u)}$ is increasing for $u \in (0,1]$. This implies that $\frac{\beta(u)}{\alpha(1/u)}$ too, is increasing for $u \in [1,\infty)$. Thus $\frac{g_1(u)}{g_2(u)}$ is an increasing function of u. Theorems 6.3.1 and 6.3.2 then give us bounds $B_1(m,n)$ and $B_4(m,n)$ respectively so that, for $0 < a \leq B(m,n) = \max(B_1(m,n), B_4(m,n))$, the estimator δ_a dominates Ψ_x.

The bounds $B_1(m,n)$ and $B_4(m,n)$ can be evaluated from the expressions given below:

$$A_1(m,n) = \frac{(n-4)(n-5)(m+5)(m+6)}{(n-2)(n-3)(m+1)(m+2)} \ ,$$

$$A_3(m,n) = \min\left\{\frac{(n-2)(n-3)}{(m+5)(m+6)}, \frac{(n-4)(n-5)}{(m+1)(m+2)}\right\}, \quad \text{and}$$

$$A_6(m,n) = \frac{(n-4)(n-5)(j+2)(j+1)\ j(j-1)(j-2)}{[(m+2)(m+1)(j+2)(j+1)\ j(j-1)(j-2) + 2(m+4)(m+2)(n-2)(n-3)(n-4)(n-5)(j+2) - 2(m+3)(m+1)(n-2)(n-3)(n-4)(n-5)(j-2)]} \ .$$

It is observed that in the evaluation of the bounds $B_1(m,n)$, $B_4(m,n)$ and Bhattacharyya's bound $A(m,n)$, $A_1(m,n)$ does not play any important role. In fact, it is seen that whenever $A_1(m,n)$ is less than 1, the bound is smaller than $2A_1(m,n)$. Thus improving $A_2(m,n)$ of Bhattacharyya's bound gives a better bound. The values of $B(m,n)$ are presented in Table 6.1 for $m = 2(1)15(5)50$ and $n = 6(1)15(5)50$. A starred entry corresponds to the case when maximum is $B_1(m,n)$. A blank entry stands for the value being 2. We observe that improvement over both Ψ_x and Ψ_y is possible for $n \geq 20$ if $m = n$ and for all $n \geq 32$ if $m \leq n+5$. Bhattacharyya through $A(m,n)$ was able to show this dominance for $n \geq 25$ if $m = n$ and for all $n \geq 35$ if $m \leq n+5$.

Since $\frac{g_1(u)}{g_2(u)}$ is increasing in u we can apply Theorems 6.4.2 and 6.4.3 also. However, as mentioned in Remark 6.4.4, Akai's (1982) result would give a larger set of c's for δ_c^* to improve Ψ_x and Ψ_y.

6.6 A Differential Inequality for Improving the Graybill-Deal Estimator of the Common Mean of Two Independent Normal Populations

Let $\underline{X} = (X_1,\ldots,X_m)$ and $\underline{Y} = (Y_1,\ldots,Y_n)$ be independent random samples from two normal populations with common mean μ and unknown and possibly unequal variances σ_1^2 and σ_2^2 respectively. The problem is to estimate μ with either of the two loss functions;

$$L_1(\hat{\mu},\mu) = (\hat{\mu}-\mu)^2 \tag{6.6.1}$$

and

$$L_2(\hat{\mu},\mu) = (\hat{\mu}-\mu)^2/\sigma_1^2 \tag{6.6.2}$$

If σ_2^2/σ_1^2 is known, the estimator $T = \alpha\bar{X} + (1-\alpha)\bar{Y}$, where

Values of B(m,n)

m	n = 6	7	8	9	10	11	12	13	14	15	20	25	30	35	40	45	50
2	.3333*	.9458	1.0082														
3	.2000*	.5832	1.1367	1.8392													
4	.1333*	.4000*	.7751	1.2674	1.6616												
5	.0952*	.2857*	.5683	.9228	1.3629	1.8785											
6	.0714*	.2143*	.4285*	.6990	1.0376	1.4361	1.8917										
7	.0556*	.1667*	.3333*	.5473	.8148	1.1389	1.4400	1.9023									
8	.0444*	.1333*	.2667*	.4444*	.6559	.9123	1.2077	1.5410	1.9110								
9	.0364*	.1091*	.2182*	.3636*	.5398	.7508	.9954	1.2720	1.5799	1.9183							
10	.0303*	.0909*	.1818*	.3030*	.4545*	.6282	.8339	1.0669	1.3267	1.6127							
11	.0256*	.0769*	.1538*	.2564*	.3846*	.5331	.7083	.9078	1.1290	1.3737							
12	.0220*	.0659*	.1319*	.2198*	.3297*	.4615*	.6000	.7803	.9719	1.1834							
13	.0190*	.0571*	.1143*	.1905*	.2857*	.4000*	.5288	.6781	.8451	1.0297							
14	.0167*	.0500*	.1000*	.1667*	.2500*	.3500*	.4667*	.5946	.7414	.9037	1.9420						
15	.0147*	.0441*	.0882*	.1471*	.2206*	.3000*	.4118*	.5255	.6555	.7993	1.7212						
20	.0087*	.0260*	.0519*	.0866*	.1299*	.1818*	.2424*	.3117*	.3896*	.4762*	1.0267	1.7821					
25	.0057*	.0171*	.0342*	.0570*	.0855*	.1197*	.1595*	.2051*	.2564*	.3134*	.6795	1.1833	1.8212				
30	.0040*	.0121*	.0242*	.0403*	.0605*	.0947*	.1129*	.1452*	.1815*	.2218*	.4839*	.8412	1.2970	1.8483			
35	.0030*	.0090*	.0180*	.0300*	.0450*	.0631*	.0841*	.1081*	.1351*	.1652*	.3604*	.6280	.9695	1.3832	1.8683		
40	.0023*	.0070*	.0139*	.0232*	.0348*	.0488*	.0650*	.0836*	.1045*	.1278*	.2787*	.4870*	.7516	1.0731	1.4506	1.8836	
45	.0019*	.0056*	.0111*	.0185*	.0278*	.0389*	.0518*	.0668*	.0833*	.1018*	.2220*	.3885*	.5994	.8564	1.1583	1.5045	1.8955
50	.0015*	.0045*	.0090*	.0151*	.0226*	.0317*	.0422*	.0543*	.0679*	.0830*	.1818*	.3167*	.4902*	.6990	.9458	1.2293	1.5493

$\alpha = \dfrac{\sigma_2^2/n}{\sigma_1^2/m + \sigma_2^2/n}$ is the best linear combination of $\bar{X}$ and $\bar{Y}$. Since T is a function of the complete sufficient statistic, it is also uniformly minimum variance unbiased estimator (UMVUE) of μ.

The problem changes completely, in case σ_2^2/σ_1^2 is unknown. In this case, the minimal sufficient statistic $(\bar{X}, \bar{Y}, S_1, S_2)$, where S_1, S_2 are the sample sums of squares, is not complete as $E(\bar{X}-\bar{Y}) = 0$. It can be shown that a UMVUE of μ does not exist. In fact, it has been established by Unni (1978) that the nonexistence of a UMVUE holds for any unbiased estimable function of μ.

One of the most commonly used estimators of μ is the Graybill-Deal (1959) estimator $\hat{\mu}_{GD}$ given by

$$\hat{\mu}_{GD} = \frac{S_2\bar{X} + S_1\bar{Y}}{S_2 + S_1} . \tag{6.6.3}$$

It was shown that $\hat{\mu}_{GD}$ dominates both $\bar{X}$ and $\bar{Y}$ if and only if m and n are at least 11. A detailed account of the results related to the properties of the Graybill-Deal estimator and its comparison with some other estimators is given in Section 1.1.2. However, so far there is no definite result proving or disproving the admissibility of $\hat{\mu}_{GD}$ in the class of all the estimators. Sinha and Mouqadem (1982) considered subclasses $\mathcal{C}$, $\mathcal{C}_o$, $\mathcal{C}_1$ and $\mathcal{C}_2$ of unbiased estimators, where

$$\begin{aligned}
\mathcal{C} &= \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \leq \Psi(S_1, S_2, (\bar{Y}-\bar{X})^2) \leq 1\} , \\
\mathcal{C}_o &= \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \leq \Psi(S_2/S_1) \leq 1\} , \\
\mathcal{C}_1 &= \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \leq \Psi(S_1, S_2) \leq 1\} , \\
\mathcal{C}_2 &= \{\hat{\mu}: \hat{\mu} = \bar{X} + (\bar{Y}-\bar{X})\Psi , \quad 0 \leq \Psi\left(\frac{S_1}{(\bar{Y}-\bar{X})^2}, \frac{S_2}{(\bar{Y}-\bar{X})^2}\right) \leq 1\} .
\end{aligned} \tag{6.6.4}$$

It was shown that $\hat{\mu}_{GD}$ is admissible in $\mathcal{C}_o$ and extended admissible in $\mathcal{C}$. Also it was shown that $\hat{\mu}_{GD}$ cannot be Bayes or limiting Bayes in $\mathcal{C}_2$, provided the prior distributions satisfy a mild condition. However, it does not prove inadmissibility of $\hat{\mu}_{GD}$.

In this section, we obtain a differential inequality, a solution to which will provide an improvement over $\hat{\mu}_{GD}$ in the class $\mathcal{C}_1$. For convenience we take $m = n$.

Let $\delta_\Psi = \overline{X} + (\overline{Y}-\overline{X})\Psi(S_1,S_2)$ be an estimator in $\mathcal{C}_1$. Writing $\underline{\theta}$ for $(\mu,\ \sigma_1^2,\ \sigma_2^2)$ we have

$$\begin{aligned} R(\underline{\theta},\delta_\Psi) &= E(\overline{X} + (\overline{Y}-\overline{X})\Psi(S_1,S_2) - \mu)^2 \\ &= E(\overline{X}-\mu)^2 + E(\overline{Y}-\overline{X})^2\Psi^2(S_1,S_2) \\ &\quad + 2E(\overline{X}-\mu)(\overline{Y}-\overline{X})\Psi(S_1,S_2) . \end{aligned}$$

Since $(\overline{X},\overline{Y})$ and (S_1,S_2) are independently distributed, we have

$$R(\underline{\theta},\delta_\Psi) = \frac{1}{n}\left[\sigma_1^2 + (\sigma_1^2 + \sigma_2^2)E\ \Psi^2(S_1,S_2) - 2\sigma_1^2\ E\ \Psi(S_1,S_2)\right] . \tag{6.6.5}$$

The estimator $\hat{\mu}_{GD}$ as defined in (6.6.3) belongs to $\mathcal{C}_1$ with $\Psi(S_1,S_2) = \dfrac{S_1}{S_1+S_2}$ for $m = n$. Denoting by T the variable $\dfrac{S_1}{S_1+S_2}$, consider

$$R(\underline{\theta},\delta_\Psi) \le R(\underline{\theta},\hat{\mu}_{GD}) \qquad \text{for all } \underline{\theta} \tag{6.6.6}$$

$$\Longleftrightarrow\ (\sigma_1^2 + \sigma_2^2)E(\Psi^2(S_1,S_2)-T^2) + 2\sigma_1^2\ E(T-\Psi(S_1,S_2)) \le 0$$

$$\text{for all } (\sigma_1^2,\ \sigma_2^2) . \tag{6.6.7}$$

Dividing both the sides by $\sigma_1^2\sigma_2^2$, the inequality (6.6.7) reduces to

$$\left(\frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}\right)E(\Psi^2(S_1,S_2)-T^2) + \frac{2}{\sigma_2^2}\ E(T-\Psi(S_1,S_2)) \le 0 \quad \text{for all } (\sigma_1^2,\ \sigma_2^2) \tag{6.6.8}$$

Stein (1973) gave an identity which gives an unbiased estimator of the risk difference of two estimators in case the underlying distribution is normal. Hudson (1978), using the same technique of integrating by parts gave the following identity for the exponential family:

Lemma 6.6.1: Let X be a continuous random variable with the density

$$f_\theta(x) = e^{\theta x - \Psi(\theta)} k(x)$$

with support (c,d), (c and d could be $-\infty$ and $+\infty$ respectively). Let

$$\lim_{x\to c} e^{\theta x} k(x) = \lim_{x\to d} e^{\theta x} k(x) = 0$$

and

$$t(x) = -\frac{k'(x)}{k(x)}\,.$$

Then for any absolutely continuous function g on R such that $E|g'(X)| < \infty$,

$$E\{(t(X)-\theta)\ g(X)\} = E\{g'(X)\} \tag{6.6.9}$$

Since $\frac{S_1}{\sigma_1^2}$ and $\frac{S_2}{\sigma_2^2}$ are independently distributed as χ^2_{n-1} random variables, the above identity can be applied. Let $h(s_1,s_2)$ be absolutely continuous in both its arguments and $E\left|\frac{\partial}{\partial S_1} h(S_1,S_2)\right|$ and $E\left|\frac{\partial}{\partial S_2} h(S_1,S_2)\right|$ be finite. Then

$$\frac{1}{\sigma_i^2} E\{h(S_1,S_2)\} = E\left[\frac{(n-3)}{S_i} h(S_1,S_2) + 2\frac{\partial}{\partial S_i} h(S_1,S_2)\right]\ ,\ i = 1,2\ . \tag{6.6.10}$$

We now obtain an unbiased estimator of the expression on the left hand side of the inequality (6.6.8) by an application of (6.6.10). The unbiased estimator turns out to be

$$Q_\Psi(S_1,S_2) = 4\Psi \frac{\partial\Psi}{\partial S_1} + 4(\Psi-1)\frac{\partial\Psi}{\partial S_2} + (n-3)\frac{(S_1+S_2)}{S_1S_2}\Psi^2$$

$$-\frac{2(n-3)}{S_2}\Psi + \frac{(n-3)S_1}{S_2(S_1+S_2)} - \frac{8S_1S_2}{(S_1+S_2)^3}. \qquad (6.6.11)$$

If we can find a Ψ^* such that $Q_{\Psi^*}(s_1,s_2) \le 0$ for all s_1, $s_2 > 0$ and $P_{(\sigma'^2_1,\sigma'^2_2)}(Q_{\Psi^*}(S_1,S_2) < 0) > 0$ for some $(\sigma'^2_1, \sigma'^2_2)$, the inequality (6.6.8) will be satisfied for all (σ_1^2, σ_2^2) with strict inequality for $(\sigma'^2_1, \sigma'^2_2)$. Thus δ_{Ψ^*} will be an improvement over $\hat{\mu}_{GD}$. We can state the above result in the following theorem.

Theorem 6.6.2: The estimator $\hat{\mu}_{GD}$ can be improved by an estimator δ_Ψ in $\mathcal{C}_1$ with respect to loss functions (6.6.1) and (6.6.2), provided Ψ satisfies

$$4\Psi\frac{\partial\Psi}{\partial s_1} + 4(\Psi-1)\frac{\partial\Psi}{\partial s_2} + (n-3)\frac{(s_1+s_2)}{s_1s_2}\Psi^2 - \frac{2(n-3)}{s_2}\Psi$$

$$+\frac{(n-3)s_1}{s_2(s_1+s_2)} - \frac{8s_1s_2}{(s_1+s_2)^3} \le 0 \qquad (6.6.12)$$

for all $s_1, s_2 > 0$ with strict inequality on a set of positive probability for some (σ_1^2, σ_2^2).

Remark 6.6.1: We do not yet have a solution for the inequality (6.6.12). However, in view of the result by Sinha and Mouqadem (1982) that $\hat{\mu}_{GD}$ is admissible in $\mathcal{C}_o$, we should search for a Ψ which does not depend on S_1, S_2 only through S_2/S_1.

6.7 Estimating the Common Mean of a Bivariate Normal Population When the Correlation is Nonzero

Let $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ be a random sample from a bivariate normal population with the common mean μ, unknown and

unequal variances σ_1^2, σ_2^2 and unknown covariance $\sigma_{12} = \rho\sigma_1\sigma_2$. Let $(\overline{X},\overline{Y})$ be the sample mean vector, $s_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$,

$s_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$ and $s_{12} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})(Y_i-\overline{Y})$. We estimate μ with the loss function

$$L_3(\hat{\mu}, a) = \frac{4n}{\sigma_1^2}(\hat{\mu}-a)^2 . \tag{6.7.1}$$

The MLE $\hat{\mu}_2$ of μ is given by

$$\hat{\mu}_2 = \overline{Y} + (\overline{X}-\overline{Y})\frac{(s_1^2-s_{12})}{(s_1^2+s_2^2-2s_{12})} \tag{6.7.2}$$

(see Halperin (1961)).

Krishnamoorthy and Rohatgi (1988a) consider

$$U_i = X_i - Y_i \quad \text{and} \quad V_i = X_i + Y_i \ , \quad i = 1,\ldots,n \tag{6.7.3}$$

and take $\overline{U}$, $\overline{V}$, s_U^2, s_V^2, s_{UV} as the sample means, variances and covariances based on (U_i, V_i). We have

$E(U_i) = 0$, $E(V_i) = 0$, $\sigma_U^2 = \mathrm{Var}(U_i) = \sigma_1^2 + \sigma_2^2 - 2\sigma_{12}$

$\sigma_V^2 = \mathrm{Var}(V_i) = \sigma_1^2 + \sigma_2^2 + 2\sigma_{12}$, $\sigma_{UV} = \mathrm{Cov}(U_i, V_i) = \sigma_1^2 - \sigma_2^2$.

In terms of the transformed variables, $\hat{\mu}_2$ can be written as

$$\hat{\mu}_2 = \tfrac{1}{2}(\overline{V} - \overline{U}\frac{s_{UV}}{s_U^2}) . \tag{6.7.4}$$

Krishnamoorthy and Rohatgi proposed an estimator $\hat{\mu}_3$, which is obtained from $\hat{\mu}_2$ by replacing s_U^2 by $\sum_{i=1}^{n} U_i^2/(n-1)$. Thus

$$\hat{\mu}_3 = \tfrac{1}{2}(\overline{V} - \overline{U}\frac{s_{UV}^*}{\sum_{i=1}^{n} U_i^2}) , \tag{6.7.5}$$

where $S^*_{UV} = (n-1)S_{UV}$.

It was shown that

$$R(\underline{\theta},\hat{\mu}_3) < R(\underline{\theta},\hat{\mu}_2) , \qquad \underline{\theta} = (\rho, \sigma_1^2, \sigma_2^2)$$

if and only if

$$R^2 < \frac{(2n-1)(n+4)}{(7n^2-8n-4)} , \tag{6.7.6}$$

where R is the correlation coefficient between U and V.

Some Modifications of $\hat{\mu}_3$

The multiple of $\frac{\bar{U}}{2}$ in $\hat{\mu}_2$ is $\frac{S_{UV}}{S_U^2}$ which is unbiased for $\frac{\sigma_{UV}}{\sigma_U^2}$. However, its modification $S^*_{UV}/\sum_{i=1}^{n} U_i^2$ in $\hat{\mu}_3$ has an expectation $(n-1)\ \sigma_{UV}/n\ \sigma_U^2$. Therefore it seems reasonable to have a multiple $\frac{n}{n-1}\frac{S^*_{UV}}{\sum_{i=1}^{n} U_i^2}$ and consider the estimator

$$\hat{\mu}_3\left(\frac{n}{n-1}\right) = \frac{1}{2}\left(\bar{V} - \frac{n}{n-1}\frac{\bar{U}\ S^*_{UV}}{\sum_{i=1}^{n} U_i^2}\right) . \tag{6.7.7}$$

The risk expression for $\hat{\mu}_3(\frac{n}{n-1})$ can be obtained from the computations in Krishnamoorthy and Rohatgi (1988a):

$$R(\underline{\theta},\hat{\mu}_3(\tfrac{n}{n-1})) = \frac{(1 + k + 2\rho\sqrt{k})}{(n-1)(n+2)(n+4)}\left[(n^2+2n-2)(n+4) - n(n^2+6n-4)R^2\right]. \tag{6.7.8}$$

We can easily prove the following result.

Theorem 6.7.1: The estimator $\hat{\mu}_3(\frac{n}{n-1})$ has risk smaller than that of $\hat{\mu}_2$ if and only if

$$R^2 < \frac{(2n-1)(n+4)}{(7n^2-12n+8)} . \tag{6.7.9}$$

Remark 6.7.1: Comparing (6.7.6) and (6.7.9), we observe that for $n > 3$, $\hat{\mu}_3(\frac{n}{n-1})$ has a larger region of dominance over $\hat{\mu}_2$ than $\hat{\mu}_3$

Next, we consider estimators $\hat{\mu}_3(c)$ proposed by Krishnamoorthy and Rohatgi (1988a),

$$\hat{\mu}_3(c) = \frac{1}{2}(\bar{V} - \frac{c\,\bar{U}\,S^*_{UV}}{\sum_{i=1}^{n} U_i^2}) , \qquad c \text{ real} . \tag{6.7.10}$$

As in the case of $\hat{\mu}_3(\frac{n}{n-1})$, the risk expression for $\hat{\mu}_3(c)$ can be obtained from the computations of Krishnamoorthy and Rohatgi:

$$R(\underline{\theta},\hat{\mu}_3(c)) = (1+k+2\rho\sqrt{k})\left[1 + \frac{c^2(n-1)}{n(n+2)} + \frac{R^2(n-1)}{(n+2)}\left\{\frac{c^2(n^2-4)}{n(n+4)} - 2c\right\}\right] . \tag{6.7.11}$$

Now

$$R(\underline{\theta},\hat{\mu}_3(c)) < R(\underline{\theta},\hat{\mu}_2)$$

if and only if

$$R^2 < B(c) ,$$

where

$$B(c) = \frac{(n+4)\left[n(n+2) - c^2(n-1)(n-3)\right]}{\left[(n-3)(n-1)\{c^2(n-2)(n+2) - 2n(n+4)c\} + n(n+2)(n+4)(n-2)\right]} . \tag{6.7.12}$$

We can get a c^* such that $\hat{\mu}_3(c^*)$ has the largest region of dominance over $\hat{\mu}_2$ among all $\hat{\mu}_3(c)$ by maximizing $B(c)$ with respect to c. The c^* can be easily seen to equal

$$\frac{(n+2)}{(n-1)}\left[\frac{(n-2)(n+3)}{(n-3)(n+4)} - \left\{\frac{6(n^3 + n^2 - 6n + 12)}{(n+2)(n-3)^2(n+4)^2}\right\}^{1/2}\right] .$$

In Table 6.2 we have presented values of c^*, $B(1)$, $B(\frac{n}{n-1})$ and $B(c^*)$ for $n = 5(5)50$. The value of c^* is seen to be larger

TABLE 6.2

Region of dominance of $\hat{\mu}_3(c^*)$, $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(1)$ over $\hat{\mu}_2$

n	c^*	$B(c^*)$	$B(\frac{n}{n-1})$	$B(1)$
5	1.299194	0.659976	0.658537	0.618321
10	1.102929	0.452566	0.452381	0.431818
15	1.058468	0.393725	0.392730	0.379738
20	1.039963	0.366106	0.364486	0.355083
25	1.030068	0.350086	0.348028	0.340635
30	1.023984	0.339629	0.337256	0.331242
35	1.019893	0.332267	0.329658	0.324569
40	1.016966	0.326803	0.324012	0.319603
45	1.014774	0.322587	0.319651	0.315753
50	1.013075	0.319235	0.316182	0.312705

than 1 and it tends to 1 as n increases to infinity. The difference $B(c^*)-B(1)$ is large compared to $B(c^*)-B(\frac{n}{n-1})$ for small values of n, and as expected, decreases to zero as n approaches infinity.

Minimal Complete Class of Estimators $\hat{\mu}_3(c)$

It is clear that $R(\underline{\theta},\hat{\mu}_3(c))$ of (6.7.11) is a convex function of c. For fixed $\underline{\theta}$, the c minimizing $R(\underline{\theta},\hat{\mu}_3(c))$ is

$$\hat{c}(\underline{\theta}) = \frac{n(n+4)R^2}{(n+4) + (n-2)(n+2)R^2}, \qquad (6.7.13)$$

which clearly depends on $\underline{\theta}$. However,

$$\inf_{\underline{\theta}} \hat{c}(\underline{\theta}) = 0 \quad \text{and} \quad \sup_{\underline{\theta}} \hat{c}(\underline{\theta}) = \frac{n+4}{n+1}.$$

An application of the Brewster-Zidek technique (1974) leads to the conclusion that the estimator $\hat{\mu}_3(c)$ for $c > \frac{n+4}{n+1}$ is improved by

$\hat{\mu}_3(\frac{n+4}{n+1})$ and $\hat{\mu}_3(c)$ for $c < 0$ is improved by $\hat{\mu}_3(0) = \frac{\bar{X}+\bar{Y}}{2}$. Also the estimators $\hat{\mu}_3(c)$, $0 \leq c \leq \frac{n+4}{n+1}$ are admissible among all $\hat{\mu}_3(c)$'s. This proves the following result.

Theorem 6.7.2: The class $\{\hat{\mu}_3(c): 0 \leq c \leq \frac{n+4}{n+1}\}$ is minimal essentially complete among the estimators of the form $\hat{\mu}_3(c)$.

Remark 6.7.2: Krishnamoorthy and Rohatgi also suggested estimators

$$\hat{\mu}_2(c) = \frac{1}{2}(\bar{V} - c\,\bar{U}\,\frac{S_{UV}}{S_U^2}) . \tag{6.7.14}$$

The c minimizing $R(\underline{\theta},\hat{\mu}_2(c))$ for fixed $\underline{\theta}$ is

$$\hat{c}(\underline{\theta}) = \frac{(n-3)R^2}{(n-4)R^2 + 1} ,$$

which attains its minimum and maximum at $R = 0$ and 1 respectively. Thus $\operatorname{Inf}_{\underline{\theta}} \hat{c}(\underline{\theta}) = 0$ and $\operatorname{Sup}_{\underline{\theta}} \hat{c}(\underline{\theta}) = 1$. Since $R(\underline{\theta},\hat{\mu}_2(c))$ is a convex function of c, we conclude once again by the Brewster-Zidek technique that the class $\{\hat{\mu}_2(c): 0 \leq c \leq 1\}$ is minimal essentially complete among the estimators of the form $\hat{\mu}_2(c)$.

Risk Comparisons

We have compared numerically the risk performances of estimators $\hat{\mu}_2$, $\hat{\mu}_3$, $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(c^*)$. We have tabulated these risks for $n = 10, 20, 40$ and various values of ρ and k (Tables 6.3 to 6.14). Although none of the estimators is seen to dominate any other estimator uniformly, some general trends can be noticed. For fixed ρ, the estimator $\hat{\mu}_2$ performs best in the region $0 < k \leq 0.4$. Similarly $\hat{\mu}_3$ performs best in $0.6 \leq k \leq 1$. However, for $0 < k \leq 0.4$, the estimator $\hat{\mu}_3$ is the worst and

Values of $A_1 = R(\underline{\theta}, \hat{\mu}_2)$, $A_2 = R(\underline{\theta}, \hat{\mu}_3)$, $A_3 = R(\underline{\theta}, \hat{\mu}_3(\frac{n}{n-1}))$ and $A_4 = R(\underline{\theta}, \hat{\mu}_3(c^*))$

$n = 10$

TABLE 6.3: $\rho = 0.2$

k	A_1	A_2	A_3	A_4
0.001	0.004440	0.094334	0.065686	0.067294
0.005	0.022466	0.111534	0.083154	0.084746
0.010	0.045243	0.132772	1.104733	0.105294
0.040	0.182857	0.257714	0.233228	0.234544
0.100	0.450799	0.498323	0.481599	0.482393
0.200	0.859565	0.864490	0.859897	0.859876
0.400	1.530428	1.467582	1.482215	1.480897
0.600	2.040941	1.930833	1.958723	1.955495
0.800	2.434327	2.292255	2.328949	2.325100
1.000	2.742857	2.680000	2.622222	2.618962

TABLE 6.4: $\rho = 0.4$

k	A_1	A_2	A_3	A_4
0.001	0.003936	0.095028	0.066000	0.067629
0.005	0.020244	0.112243	0.082869	0.084512
0.010	0.041290	0.132934	0.103599	0.105234
0.040	0.174545	0.257688	0.230592	0.232058
0.100	0.453355	0.511821	0.491603	0.492592
0.200	0.911866	0.925571	0.917997	0.918127
0.400	1.718052	1.651996	1.666991	1.665592
0.600	2.350247	2.225273	2.256810	2.254277
0.800	2.832751	2.667850	2.710407	2.707098
1.000	3.200000	3.010000	3.059259	3.055455

TABLE 6.5: $\rho = 0.6$

k	A_1	A_2	A_3	A_4
0.001	0.003038	0.095883	0.065959	0.067611
0.005	0.015898	0.111020	0.080666	0.082367
0.010	0.032873	0.129245	0.098434	0.100154
0.040	0.145286	0.240457	0.209947	0.211614
0.100	0.406052	0.482316	0.456596	0.457909
0.200	0.882111	0.915879	0.902020	0.902513
0.400	1.825567	1.767314	1.779446	1.778173
0.600	2.618151	2.484005	2.157521	2.514790
0.800	3.220878	3.034553	3.082568	3.078827
1.000	3.657143	3.440000	3.496296	3.491949

TABLE 6.6: $\rho = 0.9$

k	A_1	A_2	A_3	A_4
0.001	0.000920	0.095251	0.065201	0.066889
0.005	0.004948	0.105364	0.073632	0.075159
0.010	0.010465	0.115276	0.081855	0.083729
0.040	0.051092	0.169067	0.131310	0.133415
0.100	0.163637	0.290174	0.249294	0.251538
0.200	0.439766	0.558314	0.519007	0.521073
0.400	1.328192	1.372210	1.353517	1.354137
0.600	2.533189	2.452222	2.469098	2.467329
0.800	3.656545	3.458232	3.508532	3.504522
1.000	4.342857	4.085000	4.151852	4.146689

n = 20

TABLE 6.7: $\rho = 0.2$

k	A_1	A_2	A_3	A_4
0.001	0.004114	0.032740	0.023397	0.025259
0.005	0.020814	0.049303	0.039998	0.041849
0.010	0.041916	0.070002	0.060820	0.062643
0.040	0.169412	0.194182	0.186029	0.187520
0.100	0.417652	0.435120	0.429247	0.430334
0.200	0.796362	0.802403	0.800101	0.800399
0.400	1.417897	1.405867	1.409205	1.408252
0.600	1.890872	1.866451	1.873643	1.871828
0.800	2.255332	2.222803	2.232482	2.230095
1.000	2.541176	2.503636	2.514833	2.512085

TABLE 6.8: $\rho = 0.4$

k	A_1	A_2	A_3	A_4
0.001	0.003645	0.032651	0.023185	0.025072
0.005	0.018755	0.048133	0.038539	0.040449
0.010	0.038254	0.067629	0.058027	0.059934
0.040	0.161711	0.189074	0.180078	0.181840
0.100	0.420020	0.440983	0.433968	0.435282
0.200	0.844817	0.853924	0.850602	0.851094
0.400	1.591725	1.579650	1.582930	1.581954
0.600	2.177434	2.149904	2.157985	2.155933
0.800	2.624460	2.585738	2.597959	2.595190
1.000	2.964706	2.920909	2.933971	2.930766

TABLE 6.9: $\rho = 0.6$

k	A_1	A_2	A_3	A_4
0.001	0.002815	0.032213	0.022619	0.024531
0.005	0.014729	0.045078	0.035169	0.037142
0.010	0.030456	0.061291	0.051217	0.053220
0.040	0.135529	0.166255	0.156172	0.158155
0.100	0.376195	0.402575	0.393811	0.395482
0.200	0.817250	0.832585	0.827243	0.828143
0.400	1.691334	1.682302	1.684548	1.683758
0.600	2.425640	2.395586	2.405061	2.402880
0.800	2.984049	2.941530	2.954167	2.951043
1.000	3.388235	3.333182	3.353110	3.349447

TABLE 6.10: $\rho = 0.9$

k	A_1	A_2	A_3	A_4
0.001	0.000852	0.030871	0.021075	0.023029
0.005	0.004584	0.036560	0.026124	0.028204
0.010	0.009695	0.043099	0.032195	0.034367
0.040	0.047336	0.085139	0.072784	0.075238
0.100	0.151605	0.192719	0.179240	0.181897
0.200	0.407430	0.447439	0.434215	0.435770
0.400	1.230531	1.251451	1.244114	1.245328
0.600	2.346925	2.334349	2.337480	2.336380
0.800	3.387681	3.343616	3.356590	3.353317
1.000	4.023529	3.964091	3.981818	3.977468

n = 40

TABLE 6.11: $\rho = 0.2$

k	A_1	A_2	A_3	A_4
0.001	0.003990	0.012172	0.009472	0.010257
0.005	0.020189	0.023342	0.025651	0.026432
0.010	0.040658	0.048708	0.046050	0.046820
0.040	0.164324	0.171507	0.169129	0.169810
0.100	0.405110	0.410360	0.408607	0.409093
0.200	0.772447	0.774667	0.773895	0.774074
0.400	1.375317	1.372759	1.373533	1.373227
0.600	1.834089	1.828286	1.830108	1.829470
0.800	2.187605	2.179708	2.182202	2.181348
1.000	2.464865	2.455714	2.458608	2.457621

TABLE 6.12: $\rho = 0.4$

k	A_1	A_2	A_3	A_4
0.001	0.003537	0.011826	0.009091	0.009886
0.005	0.018192	0.026598	0.023824	0.024629
0.010	0.037105	0.045522	0.042743	0.043649
0.040	0.156855	0.164773	0.162153	0.162906
0.100	0.407407	0.413657	0.411574	0.412156
0.200	0.819447	0.822573	0.821500	0.821763
0.400	1.543925	1.541462	1.542197	1.541892
0.600	2.112046	2.105530	2.107572	2.106854
0.800	2.545648	2.536497	2.539386	2.538396
1.000	2.875675	2.865000	2.868375	2.8672[illegible]

TABLE 6.13: $\rho = 0.6$

k	A_1	A_2	A_3	A_4
0.001	0.002730	0.014434	0.008360	0.009165
0.005	0.014287	0.022967	0.020103	0.020935
0.010	0.029541	0.038370	0.035456	0.036802
0.040	0.134459	0.140322	0.137392	0.138236
0.100	0.364898	0.372668	0.370086	0.370816
0.200	0.792708	0.797596	0.795943	0.795877
0.400	1.640543	1.689011	1.639433	1.639213
0.600	2.352798	2.345001	2.348124	2.347367
0.800	2.894438	2.884139	2.887390	2.885273
1.000	3.286486	3.274286	3.278144	3.275828

TABLE 6.14: $\rho = 0.9$

k	A_1	A_2	A_3	A_4
0.001	0.000827	0.009404	0.006574	0.007397
0.005	0.004446	0.013585	0.010570	0.011447
0.010	0.009404	0.018954	0.015803	0.015719
0.040	0.045914	0.056744	0.053169	0.054206
0.100	0.147053	0.158893	0.154979	0.156109
0.200	0.395195	0.406879	0.403004	0.404107
0.400	1.193578	1.200348	1.198034	1.198627
0.600	2.276447	2.274309	2.274899	2.274592
0.800	3.285949	3.275459	3.278753	3.277601
1.000	3.902703	3.888214	3.892796	3.891233

percentage risk differences of $\hat{\mu}_3$ with $\hat{\mu}_2$ are much larger when compared to the corresponding percentage risk differences of $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(c^*)$ with $\hat{\mu}_2$. For $0.6 \leq k \leq 1$, $\hat{\mu}_2$ is the worst and the percentage risk differences of the estimators $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(c^*)$ with $\hat{\mu}_3$ are smaller compared to those of $\hat{\mu}_2$ with $\hat{\mu}_3$. Thus, $\hat{\mu}_3(\frac{n}{n-1})$ and $\hat{\mu}_3(c^*)$ are close to the best estimator in whole of the region $0 < k \leq 1$.

Similar trends are noticed for negative values of ρ, values of $k > 1$ and $n = 30$ and 50.

In view of $B(1) < B(\frac{n}{n-1}) < B(c^*)$ and the above risk observation, the estimators $\hat{\mu}_3(c^*)$ and $\hat{\mu}_3(\frac{n}{n-1})$ are good choices among $\hat{\mu}_3(c)$. The estimator $\hat{\mu}_3(\frac{n}{n-1})$ is, however, simpler to compute.

Remark 6.7.3: The risk values of $\hat{\mu}_3$ as presented in Table I of Krishnamoorthy and Rohatgi (1988a) corresponding to $\rho = 0.2$, $k = 0.1$, 0.2 and 0.4; $\rho = 0.4$, $k = 0.1$ and 0.2; $\rho = 0.6$, $k = 0.1$ and 0.2 and $\rho = 0.9$, $k = 0.1$, 0.2 and 0.4 are incorrect; the inequality (6.7.6) implies that $R(\underline{\theta},\hat{\mu}_3) > R(\underline{\theta},\hat{\mu}_2)$ for these values of ρ and k.

References

Akai, T. (1982). A combined estimator of a common parameter. Keio Science and Technology Rep., 35, 5, 93-104.

Barlow, R.E., Bartholomew, D.J., Bremner, J.M. and Brunk, H.D. (1972). Statistical Inference Under Order Restrictions, Wiley, New York.

Berger, J.O. (1985). Statistical Decision Theory and Bayesian Analysis, Springer-Verlag, New York.

Berger, J.O. and Srinivasan, C. (1978). Generalized Bayes estimators in multivariate problems. Ann. Statist. 6, 783-801.

Bhattacharyya, C.G. (1979). A note on estimating the common mean of k normal populations. Sankhya, B, 40, 272-275.

Bhattacharyya, C.G. (1980). Estimation of a common mean and recovery of interblock information. Ann. Statist. 8, 205-211.

Bhattacharyya, C.G. (1981). Estimation of a common location. Commun. Statist. - Theor. Meth. A10(10), 955-961.

Bhattacharyya, C.G. (1984). Two inequalities with an application. Ann. Inst. Statist. Math. 36, A, 129-134.

Bickel, P.J. (1981). Minimax estimation of the mean of a normal distribution when the parameter space is restricted. Ann. Statist. 9, 1301-1309.

Blackwell, D. and Girshick, M.A. (1954). Theory of Games and Statistical Decision Functions, Wiley, New York.

Blum, J.R. and Rosenblatt, J. (1967). On partial a priori information in statistical inference. Ann. Math. Statist. 38, 1671-1678.

Blumenthal, S. and Cohen, A. (1968a). Estimation of the larger translation parameter. Ann. Math. Statist. 39, 502-516.

Blumenthal, S. and Cohen, A. (1968b). Estimation of two ordered translation parameters. Ann. Math. Statist. 39, 517-530.

Brewster, J.F. and Zidek, J.V. (1974). Improving on equivariant estimators. Ann. Statist. 2, 21-38.

Brown, L.D. (1966). On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Statist. 37, 1087-1136.

Brown, L.D. (1971). Admissible estimators, recurrent diffusions, and insoluble boundary value problems. Ann. Math. Statist. 42, 855-903.

Brown, L.D. (1979). A heuristic method for determining admissibility of estimators - with applications. Ann. Statist. 7, 960-994.

Brown, L.D. and Cohen, A. (1974). Point and confidence estimation of a common mean and recovery of interblock information. Ann. Statist. 2, 963-976.

Caridi, F. (1983). Characterization of limits of Bayes procedures. J. Multivar. Anal. 13, 52-66.

Casella, G. and Strawderman, W. (1981). Estimating a bounded normal mean. Ann. Statist. 9, 868-876.

Cohen, A. (1976). Combining estimates of location. J. Amer. Statist. Assoc. 71, 172-175.

Cohen, A. and Sackrowitz, H. (1970). Estimation of the last mean of a monotone sequence. Ann. Math. Statist. 41, 2021-2034.

Cohen, A. and Sackrowitz, H. (1974). On estimating the common mean of two normal distributions. Ann. Statist. 2, 1274-1282.

Dudewicz, E.J. and Koo, J.O. (1982). Complete Categorized Guide to Statistical Selection and Ranking Procedures, American Sciences Inc., Columbus, Ohio.

Eaton, M.L. (1966). An expository presentation of a minimax theorem due to J. Kiefer. Multivar. Seminar.

Efron, B. and Morris, C. (1971). Limiting the risk of Bayes and empirical Bayes estimators, Part I: The Bayes case. J. Amer. Statist. Assoc. 66, 807-815.

Efron, B. and Morris, C. (1972a). Limiting the risk of Bayes and empirical Bayes estimators, Part II: The empirical Bayes case. J. Amer. Statist. Assoc. 67, 130-139.

Efron, B. and Morris, C. (1972b). Empirical Bayes on vector observation: An extension of Stein's method. Biometrika 59, 335-347.

Efron, B. and Morris, C. (1973a). Stein's estimation rule and its competitors - An empirical Bayes approach. J. Amer. Statist. Assoc. 68, 117-130.

Farrell, R. (1964). Estimators of a location parameter in the absolutely continuous case. Ann. Math. Statist. 35, 949-998.

Ferguson, T.S. (1967). Mathematical Statistics, A Decision Theoretic Approach, Academic Press, New York.

Gatsonis, C., MacGibbon, B. and Strawderman, W. (1987). On the estimation of a restricted normal mean. Statist. and Prob. Letters 6, 21-30.

Ghosh, M.N. (1964). Uniform approximation of minimax point estimates. Ann. Math. Statist. 35, 1030-1047.

Glesser, L.J. (1987a). Improved estimators of mean response in simulation when control variates are used. Tech. Rep., Mimeo Series, No. 87-32, Dept. of Statist., Purdue Univ.

Glesser, L.J. (1987b). Improved estimators of mean response in simulation when control variates with unknown covariance matrix are used. Tech. Rep. No. 87-34, Dept. of Statist., Purdue Univ.

Graybill, F.A. and Deal, R.B. (1959). Combining unbiased estimators Biometrics 15, 543-550.

Gupta, A.K. and Rohatgi, V.K. (1980). On the estimation of restricted mean. J. Statist. Planning and Inference 4, 369-379.

Halmos, P.R. (1952). Measure Theory, Springer International Student Ed.

Halperin, M. (1961). Almost linearly-optimum combination of unbiased estimates. J. Amer. Statist. Assoc. 56, 36-43.

Hammersley, J.M. (1950). On estimating restricted parameters. J. Roy. Statist. Soc., Ser. B, 12, 192-229.

Hogg, R.V. (1960a). Certain uncorrelated statistics. J. Amer. Statist. Assoc. 55, 265-267.

Hogg, R.V. (1960b). On conditional expectations of location statistics. J. Amer. Statist. Assoc. 55, 714-717.

Horvitz, D.G., Greenberg, B.G. and Abernathy, J.R. (1976). Randomized response: A data-gathering device for sensitive questions. Inf. Statist. Rev. 44, 181-196.

Huber, P.J. (1974). Fisher information and spline interpolation. Ann. Statist. 2, 1029-1034.

Hudson, H.M. (1978). A natural identity for exponential families with applications in multiparameter estimation. Ann. Statist. 6, 473-484.

James, W. and Stein, C. (1961). Estimation with quadratic loss. Proc. Fourth Berkeley Symp. Math. Statist. Prob. 1, 361-379.

Karlin, S. (1958). Admissibility for estimation with quadratic loss. Ann. Math. Statist. 29, 406-436.

Katz, M.W. (1961). Admissible and minimax estimates of parameters in truncated spaces. Ann. Math. Statist. 32, 136-142.

Katz, M.W. (1963). Estimating ordered probabilities. Ann. Math. Statist. 34, 967-972.

Khatri, C.G. and Shah, K.R. (1974). Estimation of location parameters from two linear models under normality. Commun. Statist. - Theor. Meth. 3, 647-663.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. Ann. Math. Statist. 28, 573-601.

Krishnamoorthy, K. and Rohatgi, V.K. (1988a). Estimation of common mean of a bivariate normal distribution. Preprint.

Krishnamoorthy, K. and Rohatgi, V.K. (1988b). Estimation of mean response when control variables are used. Preprint.

Kubokawa, T. (1987a). Estimation of a common mean of two normal distributions. Tsukuba J. Math. 11, 157-175.

Kubokawa, T. (1987b). Admissible minimax estimation of a common mean of two normal populations. Ann. Statist. 15, 1245-1256.

Kubokawa, T. (1987c). Estimation of a common mean with symmetric loss. J. Japan Statist. Soc. 17, 75-79.

Kubokawa, T. (1988). Inadmissibility of truncated estimators in balanced incomplete block designs. J. Japan Statist. Soc. 18, 1, to appear.

Kumar, S. and Sharma, D. (1988). Simultaneous estimation of ordered parameters. Commun. Statist. - Theor. Meth. 17, 12, to appear.

LeCam, L. (1955). An extension of Wald's theory of statistical decision functions. Ann. Math. Statist. 26, 69-81.

Lehmann, E.L. (1959). Testing Statistical Hypothesis, Wiley, New York.

Lehmann, E.L. (1983). Theory of Point Estimation, Wiley, New York.

Mehta, J.S. and Gurland, J. (1969). Combinations of unbiased estimators of the mean which consider inequality of unknown variances. J. Amer.Statist. Assoc. 64, 1042-1055.

Melkman, A.A. and Ritov, Y. (1987). Minimax estimation of the mean of a general distribution when the parameter space is restricted. Ann. Statist. 15, 432-442.

Moors, J.J.A. (1981). Inadmissibility of linearly invariant estimators in truncated parameter spaces. J. Amer. Statist. Assoc. 76, 910-915.

Moors, J.J.A. (1985). Estimation in truncated parameter spaces. Ph.D. dissertation, Tilburg Univ., The Netherlands.

Moors, J.J.A. and Van Houwelingen, J.C. (1987). Estimation of linear models with inequality restrictions. Res. Memo. Dept. of Economics, Tilburg Univ., The Netherlands.

Nair, K.R. (1982). An estimator of the common mean of two normal populations. J. Statist. Planning and Inference, 6, 119-122.

Norwood, T.E. and Hinkelmann, K. (1977). Estimating the common mean of several normal populations. Ann. Statist. 5, 1047-1050.

Rafsky, L.C. (1976). Inadmissibility of the observed proportion in large samples drawn with replacement. Rep. Bell. Laboratories.

Rastogi, S.C. and Rohatgi, V.K. (1974). Unbiased estimation of the common mean of a bivariate normal distribution. Biom. Z. 16, 155-166.

Sackrowitz, H. (1970). Estimation for monotone parameter sequences: the discrete case. Ann. Math. Statist. 41, 609-620.

Sackrowitz, H. (1982). Procedures for improving upon the MLE of ordered binomial parameters. J. Statist. Planning and Inference 6, 287-296.

Sackrowitz, H. and Strawderman, W. (1974). On the admissibility of the MLE for ordered binomial parameters. Ann. Statist. 2, 822-828.

Sacks, J. (1963). Generalized Bayes solutions in estimation problems. Ann. Math. Statist. 34, 751-768.

Savage, L.J. (1954). The Foundations of Statistics, Wiley, New York.

Schafer, R.E. (1976). Modified estimators for the binomial parameter. Amer. Statist. 30, 98-100.

Sharma, D. (1983). Sufficiency of a sufficient statistic for equivariant estimation. Commun. Statist. - Theor. Meth. 12, 611-624.

Shinozaki, N. (1978). A note on estimating the common mean of K normal distributions and the Stein problem. Commun. Statist. - Theor. Meth. A7(15), 1421-1432.

Sinha, B.K. and Mouqadem, O. (1982). Estimation of the common mean of two univariate normal populations. Commun. Statist. - Theor. Meth. A11, 1603-1614.

Skibinsky, M. and Cote, L. (1963). On the admissibility of some standard estimates in the presence of prior information. Ann. Math. Statist. 34, 539-548.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. Proc. Third Berkeley Symp. Math. Statist. Prob. 1, 197-206.

Stein, C. (1973). Estimation of the mean of a multivariate normal distribution. Proc. Prague Symp. on Asymptotic Statistics, 345-381.

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. Ann. Statist. 9, 1135-1151.

Unni, K. (1978). The theory of estimation in algebraic and analytic exponential families with applications to variance components models. Ph.D. dissertation, Indian Statistical Institute, Calcutta.

Wijsman, R.A. (1976). Correction to the paper "On estimating the common mean of two normal distributions" by Cohen and Sackrowitz, Ann. Statist. 4, 1294.

Zacks, S. (1966). Unbiased estimation of the common mean. J. Amer. Statist. Assoc. 61, 467-476.

Zacks, S. (1970). Bayes and fiducial equivariant estimators of the common mean of two normal distributions. Ann. Math. Statist. 41, 59-69.